

R E P O R T R E S U M E S

ED 012 380

AA 000 132

BASES OF SCHOOL CHILDREN'S ESTHETIC JUDGMENT AND ESTHETIC PREFERENCE.

BY- CHILD, IRVIN L. AND OTHERS
YALE UNIV., NEW HAVEN, CONN.

REPORT NUMBER CRP-2840

REPORT NUMBER BR-5-0205

PUB DATE JUN 67

CONTRACT OEC-5-10-048

EDRS PRICE MF-\$0.75 HC-\$6.72 168P.

DESCRIPTORS- *CHILD DEVELOPMENT, *ART APPRECIATION, *PERSONAL VALUES, *CULTURAL FACTORS, *CULTURAL ENRICHMENT, NEW HAVEN

RESEARCH INTO CHILDREN'S ART PREFERENCES AND THEIR RELATION TO THE ESTHETIC MERIT OF ART AS JUDGED BY EXPERTS WAS DESCRIBED IN THIS REPORT THAT IS MADE UP OF NINE ARTICLES, ORIGINALLY PREPARED FOR PUBLICATION IN SCHOLARLY JOURNALS, AND A SUMMARY OF THE FINDINGS. THE GENERAL TOPICS CONSIDERED WERE (1) REASONS CHILDREN GIVE FOR THEIR PREFERENCES, (2) ASSESSMENT OF SPECIFIC ASPECTS OF ESTHETIC VALUE IN VISUAL ART, (3) DETERMINANTS OF CHILDREN'S ART PREFERENCES AND OF THEIR AGREEMENT WITH EXPERT ESTHETIC JUDGEMENT (SENSORY QUALITIES, EMOTIONAL QUALITIES, ABUNDANCE, SIMPLIFICATION), (4) AGE AND SEX DIFFERENCES IN CHILDREN'S COLOR PREFERENCES, AND (5) transcultural comparisons of ESTHETIC JUDGEMENT AND ART PREFERENCE. FROM THE FINDINGS OF THE STUDIES, THE INVESTIGATORS CONCLUDED THERE WAS EVIDENCE THAT (1) ESTHETIC APPRECIATION IS AT LEAST PARTLY BASED ON HUMAN UNIVERSALS RATHER THAN ON ARBITRARY TRADITIONS OF THE CULTURALLY ELITE, AND (2) AN ESTHETIC APPROACH APPEARS NOT TO BE CONGENIAL TO MANY. PERHAPS MOST, PEOPLE. THE AUTHORS SUGGESTED THAT ART EDUCATION IN GENERAL SCHOOLING SHOULD GIVE OPPORTUNITY AND ENCOURAGEMENT FOR DEVELOPMENT OF ESTHETIC APPRECIATION, BUT AT THE SAME TIME SHOULD OFFER OPPORTUNITY FOR THE DEVELOPMENT OF VERY DIFFERENT APPROACHES TO ART. (AL)

ED012380

BR-5-0205

P.A. 24

FINAL REPORT
Project No. 2840
Contract No. OE 5-10-048

**BASES OF SCHOOL CHILDREN'S ESTHETIC JUDGMENT
AND ESTHETIC PREFERENCE**

June 1967

**U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE**

**Office of Education
Bureau of Research**

AA000132

Bases of School Children's Esthetic Judgment and Esthetic Preference

Project No. 2840
Contract No. OE 5-10-048

Irvin L. Child and others

June, 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Yale University

New Haven, Connecticut

CONTENTS

ACKNOWLEDGMENTS	iii
SUMMARY	1
REASONS CHILDREN GIVE FOR THEIR ART PREFERENCES: RELATION TO AGE, SEX, AND ESTHETIC APPROACH (Irvin L. Child & Rosaline S. Schwartz)	11
AN ATTEMPT TO ASSESS SPECIFIC ASPECTS OF ESTHETIC VALUE IN VISUAL ART (Irvin L. Child & Rosaline S. Schwartz)	54
DETERMINANTS OF CHILDREN'S ART PREFERENCES AND OF THEIR AGREEMENT WITH EXPERT ESTHETIC JUDGMENT (Irvin L. Child)	67
AGE AND SEX DIFFERENCES IN CHILDREN'S COLOR PREFERENCES (Irvin L. Child, Jens A. Hansen, & Frederick W. Hornbeck)	89
BAKWELE AND AMERICAN ESTHETIC EVALUATIONS COMPARED (Irvin L. Child & Leon Siroto, reprinted from <u>Ethnology</u> , 1965, 4, 349-360, by permission of G. P. Murdock. Not copyrighted.)	117
SOME TRANSCULTURAL COMPARISONS OF ESTHETIC JUDGMENT (C. S. Ford, E. Terry Prothro, & Irvin L. Child, reprinted from <u>The Journal of Social Psychology</u> , 1966, 68, 19-26, by permission of The Journal Press. Copyrighted 1966.)	129
COMPARISON OF ESTHETIC JUDGMENTS BY AMERICAN EXPERTS AND BY JAPANESE POTTERS (Sumiko Iwao & Irvin L. Child, reprinted from <u>The Journal of Social Psychology</u> , 1966, 68, 27-33, by permission of The Journal Press. Copyrighted 1966.)	137
FURTHER EVIDENCE OF AGREEMENT BETWEEN JAPANESE AND AMERICAN ESTHETIC EVALUATIONS (Sumiko Iwao, Irvin L. Child, & Miguel García)	144
PERSONALITY AND ESTHETIC SENSITIVITY: EXTENSION OF FINDINGS TO YOUNGER AGE AND TO DIFFERENT CULTURE (Irvin L. Child & Sumiko Iwao)	151

(The page numbers here appear in the lower corners of the pages. Each article also has separate pagination.)

ACKNOWLEDGMENTS

We prefer not to identify the four communities whose school children participated in this research. Without naming them, however, we wish here to express our profound gratitude for their invaluable aid, indispensable to the research.

BASES OF SCHOOL CHILDREN'S ESTHETIC JUDGMENT AND ESTHETIC PREFERENCE

Research performed under Contract No. OE 5-10-048 (Project No. 2840) with the Office of Education, U. S. Department of Health, Education, and Welfare.

Irvin L. Child et al.

Yale University

SUMMARY OF FINAL REPORT

This report consists of nine articles describing research into children's art preferences and their relation to esthetic merit of art (as judged by experts), to objective characteristics of art, and to personality, and research into transcultural aspects of art judgment. Each of the first four articles deals with a distinct problem and is therefore considered under a separate heading in this summary; the last five articles are mainly concerned with one general problem and are discussed in this summary under a single heading.

I. Reasons children give for their art preferences (Irvin L. Child and Rosaline S. Schwartz)

We had previously shown to large groups of children numerous pairs of projected slides and asked them to indicate their preference in each pair. A pair consists of two works of art, similar in type or subject matter and usually similar in style, differing in esthetic merit in the opinion of the person--ordinarily a present or former university art student--who made up the pair. The direction of the difference had been confirmed by 12 or more out of 14 experts--again, mostly present or former university art students--who were asked to judge independently which work in each pair was esthetically better.

After scoring each child's preferences for percentage of agreement with experts' judgments on the pairs, we used the scores to select 124 children for further study. These children individually were shown a standard set of 30 pairs and requested to indicate and then justify a preference in each pair. We interviewed in each grade from first through twelfth (corresponding to typical ages from six through seventeen) not a representative sample but two extremes within each sex: high-scorers, who tend to prefer art that experts consider esthetically better, and low-scorers, who tend to have contrary preferences.

We find marked differences between reasons given by older and younger children, and we find marked differences between reasons given by high-scorers and low-scorers. Between the two variables--age and scoring group--there are similarities. Some reasons offered frequently by older children are also offered by high-scorers, and some reasons often given by younger children are also often given by low-scorers. Yet on the whole the two variables are decidedly different. The difference appears when we try to characterize the set of reasons showing significant and especially close dependence upon age and to compare this with the set showing comparable dependence upon scoring group. With very good consistency, the two sets form distinct patterns.

Younger children, more than older ones, are apt to mention subject matter. Younger children are also much more likely to mention the evocativeness of art, or to imply it by describing a story or dramatic setting the art suggests. The decline with age in use of these two criteria is probably a consequence of increasing capacity for objectivity. Favored by young children among the specific reasons we have classified as references to inherent characteristics of art are colorfulness and presence of many things. These are both simple bases for evaluation: the conspicuously stimulating, the conspicuously ample. Younger children tend also to give

what we have called sentimental reasons for choice--liking the pretty, familiar, and pleasant, for example. Sentimental reasons may be considered part of the relative simplicity of the younger children in their interaction with art; accentuating the positive may serve to ward off the danger of emotion in the simplest possible way.

Older children's reasons for choice show greater differentiation. Their statements about color, though often vague, imply attention to complex relations, and their reasons referring to other inherent characteristics tend to be more complex than the quantitative criterion common in younger children. Yet the way the older ones allude to some of these more complex features seems to imply a continuing wish for simplicity. This inference is supported by their citing symmetry, unity, composition, and ease of understanding as reasons for choice. Additional evidence of older children's differentiated awareness may be found in the variety of reasons they give referring to objective characteristics of art--softness, strength, smoothness, roughness, and texture generally. (These reasons might also be indicative of interest in possible symbols of masculinity and femininity.) Older children tend to give fewer sentimental reasons than the younger. To some extent emotional reasons replace the sentimental, but this shift is more closely associated with scoring group than with age, and the principal correlate of age here is the decline, with years, in sentimentality.

Scoring groups, in their comments about color and light, seem to differ in very much the same way we have described age groups as differing. The low-scorers are more inclined to refer to simple colorfulness, the high-scorers to something more complex (in this instance, the blending of colors). For other inherent characteristics, however, the pattern of distinction between scoring groups is very different from that between age groups, though some of the correlations are similar. The low-scorers say they like clarity and detail; most of the high-scorers' reasons in this category indicate a liking for lack of detail and order. This pattern--low-scorers preferring the clear, direct, and simple, high-scorers liking haziness and disorder--is repeated and strengthened in the high-scorers' expressions of liking activity, roughness, texture, depth, and challenges to understanding. The pattern is a familiar one in recent research on personality in relation to creativity and esthetic responsiveness.

Emotional reasons given by high-scorers also coincide with this pattern. Liking art with more feeling, with more faithfulness to the spirit of reality and yet with emotionally meaningful stylization rather than photographic literalness, is clearly a better fit with the high-scorer's liking disorder and lack of detail than it is with the older child's liking unity and balance.

One set of reasons correlated with scoring groups forms a quite different pattern. Liking art that seems old or valuable, or art that is recognized, are reasons requiring knowledge of art. High-scorers certainly do seem to know much more about art. But all the reasons they give may not represent what they have absorbed from the cultural stream of words

about art. Symmetry, unity, and composition are such words, yet the high-scorers show no special tendency to use them; it is our older children as a whole who do, not high-scorers in particular.

Despite considerable agreement between the correlates of age and the correlates of scoring group, then, we conclude that between the two there are important distinctions which we may characterize in this way:

(1) Young children like simple and ample stimulation which is kept attractive and unthreatening. As they grow older they can tolerate emotional discomfort and can accept complexity; but they like the complexity to be unified, balanced, and easily understood.

(2) Children (regardless of age) who have an esthetic approach like disorder, lack of detail, and various complexities of form and emotion which pose a challenge to understanding, whereas children who lack an esthetic approach like clarity and detail.

Unlike age and esthetic approach, the variable of sex seems almost entirely irrelevant to the justification of art preferences. If future studies of response to art continue to confirm the near-absence of sex differences, this will be an important point to take account of in a general psychological theory of art.

II. An attempt to assess specific aspects of esthetic value in visual art (Irvin L. Child and Rosaline S. Schwartz)

Variation in preference with age and with esthetic approach might be related to specific aspects of the esthetic merit of the art concerned. To study this question, it would be necessary first to measure these. We developed a manual of instructions for rating 14 aspects of esthetic merit and were able to get two qualified judges to apply it to rating the contrasts in each of our 900 pairs of art works.

The two judges always showed some agreement on specific aspects; interjudge correlation ranged from .08 to .33. Much of this agreement, however, actually had to do with general esthetic merit. Each judge's ratings of specific aspects were heavily loaded on a general merit factor, and the two judges' general factors showed some agreement. When the general factor was statistically removed, the remaining agreement on specific ratings was very low.

The result does not give much hope that the ratings we obtained from just two judges can be usefully related to other facts about the pairs. But it strongly suggests that where comparable ratings can be obtained from many skilled judges, they can be useful in research.

III. Determinants of children's art preferences and of their agreement with expert esthetic judgment (Irvin L. Child)

To what extent do the children who have responded to our paired slides prefer in each pair the work experts consider to be better art? Our answer is based on preferences in our 900 pairs of slides expressed by several thousand Connecticut children representing the 12 public-school grades. In the elementary schools, agreement with experts averages about 40%, indicating a definite tendency to prefer the work the experts consider poorer. Through the elementary-school years there is no consistent change in the percentage. In the six grades of secondary school, on the other hand, we do find regular change; the percentage of agreement rises year by year until it finally approaches 50%. At every age this percentage varies widely from one pair to another, and the change of percentage with increasing age also varies greatly among pairs in rate and even direction. The percentage often differs very much too between two groups of secondary-school children isolated for study: high-scorers (whose preferences tend to agree with expert judgment) and low-scorers (whose preferences tend to disagree with expert judgment). We tried to learn more about children's art preferences by studying how these variations are related to characteristics of the art.

Characteristics discussed here are objective rather than evaluative. That is, they do not pertain directly to the relative esthetic merit of the two works and can be judged without reference to esthetic merit. For each of our 900 pairs, two adults assessed which of the two works was higher on each of 24 objective characteristics, and by how much. Seventeen of the variables yielded relevant results. We have grouped them under four headings:

1. Sensory qualities. Five out of six sensory qualities show decidedly uniform results. Their correlations with elementary-school preferences are all positive; at that age, children tend to prefer the picture with more light-dark contrast, warmer colors, more variation of hue, more saturated colors, and greater variation in saturation. We could generalize that the younger children prefer more colorful art. In the high-school students this consistency has disappeared: these sensory variables have little relation to their preferences. The shift is expressed also by a uniformly negative correlation between these variables and the change in preference from elementary school to high school. When we consider separately the low-scorers and high-scorers in the secondary schools, we find that the low-scorers still share some of the younger children's preference for colorfulness. The high-scorers, on the other hand, show definite correlations in the other direction. Changes in sensory preferences, except for light-dark contrast, are relevant to agreement with expert judgment, for the works experts consider better are on the average less colorful.

2. Emotional qualities. Here are grouped happiness, sentimentality, and emotionality. Happiness has an important positive influence on the preferences of elementary-school children. The low-scorers in secondary school are still about as much influenced by happiness as are the younger children; but in the high-scorers any correlation of preference with it has completely disappeared. The differentiation between better and poorer work

is more consistent on this characteristic than on any of the other 23. Results for sentimentality are similar though not so extreme. Those for emotionality tend to be the reverse, but little can be said with confidence about the results for emotionality because on this one variable interjudge reliability was very poor.

3. Abundance. Under this heading are grouped completeness, abundance of skill (difficulty of making), and abundance of detail (quantitative fullness). Each of them strongly influences the preferences of elementary-school children. In secondary school, the low-scorers still exhibit preferences almost as strongly determined by these variables as are the preferences of the elementary-school population; on the high-scorers, however, these variables have greatly diminished influence. Despite the generally predominant influence of these variables, then, some of the secondary-school students are becoming free of it. Differentiation within the pairs is in a uniform direction, the poorer work being higher in abundance.

4. Simplification. Here are grouped realism, conventionality, clarity, lack of ambiguity, and sharpness of edges--which seem to have in common that preferring them suggests a desire to make or to keep experience simple. For realism of representation, results are somewhat distinctive in that its general influence seems as great in high as in elementary school. The other four variables yield very homogeneous results. They are highly related to preference in elementary-school children and only a little less so in high-school children. Change in preference from the earlier age to the later is substantially correlated with these kinds of differentiation; the more a work is distinguished from its mate in any of these aspects of simplicity, the more is its choice likely to decrease from elementary-school to high-school years. The low-scorers in secondary school are even more influenced by these five characteristics than is the total population in elementary school; in the high-scorers, on the other hand, their influence is much reduced. Finally, all of these variables tend to be lower in the better work than in the poorer, so that these changes are part of an increasing tendency to choose works considered better.

The theme most appropriate to emphasize in over-all interpretation of these results is simplicity vs. complexity. Art's esthetic function seems to involve the challenge and the mastery of complexity. But younger children usually cannot absorb or accept complexity. They like art in which they find the security of stability, of living in a simple world they can understand. They value art that affirms happiness or gives occasion for other, even unhappy, emotions in a setting which offers bland simplicity, containment, and only superficial disturbance. This blandness can be relieved by the basically pleasant stimulation of warm, saturated colors and of color variation. They value, too, art that suggests abundance through its wealth of content or through the skill exerted to produce it, and here also is security and the simplicity of completeness. As children grow older, some of them begin to tolerate complexity and even to seek it out. They look for more challenging emotional experience in art and will tolerate incompleteness and lack of abundance. The sensory stimulation of strong color is no longer necessary and may even be a distraction.

The article reporting these findings is a preliminary one prepared for presentation at an international colloquium in September, 1966, and to be published in Italy. A final version for publication in an American journal will include additional findings and will relate the findings to those of the interview study appearing as the first article in this report.

IV. Age and sex differences in children's color preferences (Irvin L. Child, Jens A. Hansen, and Frederick W. Hornbeck)

In seeking to understand children's art preferences, a promising line of inquiry is to test whether the basis of these preferences extends beyond the realm of art to all relevant stimuli. We had originally planned to study whether specific aspects of esthetic evaluation relevant to children's art preferences could be paralleled in simple experimental stimuli and would be found there also to determine preference. Our failure to obtain good interjudge agreement on specific aspects of esthetic evaluation prevented us from carrying out that plan within the limited resources of this project.

We concentrated instead on objective characteristics, and specifically on the sensory qualities most genuinely isolable in simple stimuli. Our inquiry had two parts. The part concerned with preference for one color over another is summarized here. The other, concerned with preference for one combination over another, is not yet finished. Discussion of the implications of the single-color inquiry for art preferences had best be held off until completion of the color-combination inquiry. Meanwhile, the single-color study is of interest for the additional information it supplies on cognitive functions and simple preferences in childhood.

Children in grades 1-12 expressed personal preference for one color rather than another, with the two colors differing on either one dimension or two. Many preferences and age and sex differences appeared specific to the pair, but some generalizations were established: general preference for cool hues and high saturation; female preference for lighter colors; and, with increasing age, a decreasing preference for high saturation, an increasing consistency of hue choices relative to saturation choices, and an increasing tendency to resolve conflicts in favor of hue rather than saturation. The last two findings, interpreted in relation to the child's increasing cognitive differentiation, lead to predictions for future test.

V. Trans-Cultural Comparisons of Esthetic Judgment and Art Preference

Interpretation of evaluative responses to art must be greatly influenced by suppositions about whether they show some degree of universal consistency or, on the other hand, show no consistency or uniformity beyond those of personality and culture. Comparing responses collected from different societies is therefore fundamental to proper understanding. In the course of the present research, it has been possible on several occasions to secure, through psychologists living abroad or anthropological field

workers interested in esthetics, data on how art-oriented people in other societies respond to some of the art we have used. Our major focus has not been on trans-cultural problems, so that this work was small in scale, and simply fitted into the time left by unavoidable delays in the main project. Because of their nearly complete novelty, however, even these fragments of trans-cultural work are significant, and the five resulting brief papers are included in this report and summarized below.

- A. BaKwele and American esthetic evaluations compared (Irvin L. Child and Leon Siroto) Published in Ethnology, 1965, Vol. 4, No. 4, 349-360.

Photographs of BaKwele masks were judged for esthetic merit by art experts in New Haven, Connecticut. During a field trip to the Congo Republic (Brazzaville), judgments of these same photographs were obtained from 16 BaKwele men interested in or knowledgeable about their masks. The consensus of the 16 BaKwele and of subgroups of them, and most of the 16 individuals, showed significant agreement with the consensus of New Haven experts. The finding of some trans-cultural agreement cannot be interpreted confidently from this one study alone; it is consistent, however, with the notion that the esthetic appeal of a work of art to an art-involved viewer is partly a function of universals of human nature, and it should encourage further trans-cultural comparison of evaluative responses to art. The BaKwele also showed agreement among themselves on other bases than those shared with the New Haveners.

- B. Some transcultural comparisons of esthetic judgment (C. S. Ford, E. Terry Prothro, and Irvin L. Child) Published in The Journal of Social Psychology, 1966, 68, 19-26.

Six residents of a remote Fijian village, all engaged to some extent in craftwork, and four craftsmen in the Cycladic Islands of Greece, expressed preferences within trios of works of art that had earlier been evaluated for relative esthetic merit by Americans greatly interested in art. Significant evidence of agreement was found between Fijians and Americans. Greek craftsmen also tended toward agreement with the American judges, but the tendency was not statistically significant. Taken together, these two sets of data strongly suggest a tendency, among people actively interested in art, for their evaluations of art to agree across cultural boundaries. Previous studies that have looked for and failed to find trans-cultural agreement in art preference between people not selected for interest in art should be re-evaluated in the light of the present data. It may be that esthetic responses to works of art are made only by some people in most communities and yet show some consistency wherever they are made.

- C. Comparison of esthetic judgments by American experts and by Japanese potters (Sumiko Iwao and Irvin L. Child) Published in The Journal of Social Psychology, 1966, 68, 27-33.

Japanese potters were asked to indicate the better work of art within each of many pairs of reproductions for which American experts had

shown excellent agreement about the direction of esthetic superiority. Thirty potters judged 21 pairs of black-and-white reproductions, another 30 potters judged 15 other black-and-white pairs, and all 60 judged 12 pairs of colored abstract paintings. Agreement with the American experts was found in 62% of the judgments on the first batch and 59% of the judgments on the second batch of black-and-white pairs, and in 57.5% of the judgments on the colored pairs. Each of these findings differs significantly from the chance value of 50%. The agreement found, since it exceeds that of many non-expert Americans, must be connected with the artistic interest or activity of the Japanese potters. It may arise partly from the independent discovery, by people in differing cultural traditions, of similar aspects of art that satisfy esthetic interests.

D. Further evidence of agreement between Japanese and American esthetic evaluations (Sumiko Iwao, Irvin L. Child, and Miguel Garcia)

Thirty-one residents of Tokyo, all practitioners or teachers of flower arranging, tea ceremony, or other traditional arts, judged which was the esthetically better work in each of 51 pairs of black-and-white photographic prints and 24 pairs of colored abstract-painting postcards. U. S. experts had shown good agreement among themselves about which work in each pair was better. The Japanese subjects showed significant tendency to agree with the U. S. experts on the prints. They did not on the abstract paintings; yet comparison with a group of adults lacking esthetic orientation showed the Japanese subjects to differ significantly from the decided tendency of the other group to prefer the works spurned by U. S. experts. These findings add to the growing evidence that esthetic evaluations developed in diverse cultural settings may tend toward agreement.

E. Personality and esthetic sensitivity: Extension of findings to younger age and to different culture (Irvin L. Child and Sumiko Iwao) To be published in the Journal of Personality and Social Psychology, 1968.

We report here two confirmations of a positive relationship, previously found in U. S. college men, of esthetic sensitivity to cognitive independence and openness. A study of U. S. secondary-school students confirms the general relationship for both sexes, a younger age, and a more diversified population, while not confirming a specific relationship with a measure of Regression in the Service of the Ego. A study of college men in Japan confirms the relationship for all measures used.

Subsidiary evidence seems to justify our rejecting the hypothesis that the trans-cultural confirmation is an artifact, an effect of Westernization on response to art and to questionnaire items. Perhaps this relation between personality and esthetic sensitivity may be found in any society where esthetic values are stressed in some generally available part of the cultural tradition--as in Japan and in Western European tradition--so that the individual with cognitive independence and openness has esthetic activity available to him as one possible medium for expression and gratification of these cognitive tendencies.

Implications for Education

From the work reported here, useful suggestions may be drawn by art educators. One general theme with possible practical implications recurs throughout the research. According to evidence presented here, evaluative approaches to art often thought limited to the Western cultural tradition are found in strikingly different societies. They are not found in everyone there. Nor are they found in everyone in our own society. The esthetic approach seems not to be congenial to everyone; certain personality characteristics seem to favor it, and these--independence of judgment, regression in the service of the ego, and tolerance of complexity, ambiguity, and unrealistic experience--are correlated with it in the two diverse societies so far sampled. To some extent these characteristics are also expressed in the justifications offered for art preferences by those U. S. school children tested whose preferences most resemble the ones of art experts. Children who reject the art which an esthetic approach would favor give very different justifications for their preferences.

In providing evidence that esthetic appreciation is at least partly based on human universals rather than on arbitrary traditions of our culturally elite, the research strengthens the argument for giving esthetic appreciation a prominent place in the aims of art education. On the other hand, it provides suggestive evidence that esthetic appreciation may not be at all suited to many--perhaps most--people. If the trend of these findings is substantiated, then, it would appear that to serve in general schooling art education needs to give opportunity and encouragement for development of esthetic appreciation, but at the same time to offer opportunity and encouragement for the development of other very different approaches to art.

REASONS CHILDREN GIVE FOR THEIR ART PREFERENCES:

RELATION TO AGE, SEX, AND ESTHETIC APPROACH

Irvin L. Child

and

Rosaline S. Schwartz

Yale University

What children prefer in art, and the reasons they give for their preferences, have been investigated by educators and psychologists since the 1920's. Some studies are concerned only with the preferences children express by their choices or rankings of art; others investigate--either exclusively or along with overt preferences--reasons children give to justify their preferences. Many studies appear to be mainly quests for solutions to immediate practical problems of art education, not clearly related to general psychological problems nor to previous research on the subject, and most have been isolated and sporadic.

Two recent papers seem to us especially valuable because of their relation to the interests of developmental psychology. Machotka (1966) has studied reasons given by children aged 6 through 12, and aged 18. Francès (1966) compares reasons given by children 6 to 12 with reasons given by children 13 to 16, and also compares both these groups with older people including groups with special artistic training.

We report here our own study of reasons children give for choice.

We did not know of Francès' work until our own was almost completed. We were familiar with Machotka's, and might have done well to base our analysis on his. We decided, however, that our purposes and methods were sufficiently different to develop our own scheme empirically from our interviews. Thus our analysis is not closely similar to either Machotka's or Francès', yet some of our findings can be related to theirs. Both Machotka and Francès studied French children. Both authors studied preferences as well as their justification, though like us, both found it useful to report on preferences separately (cf. Machotka, 1963; Francès & Voillaume, 1964). Machotka's preference study embraced U. S. subjects as well as French, and he discusses some points of possible cultural difference as well as agreement. On the points where we can make comparisons with these other two studies, cultural differences do not seem especially relevant. We will be dealing with developmental trends which it is interesting, but not surprising, to find in similar form in two industrial nations with common historical roots and current cultural interaction.

An important distinction between our work and the others' lies in the fact that our selection of pictures to be commented on was directed at studying esthetic value psychologically. We used pairs of slides of art, the two in each pair being always similar in subject matter or kind and usually similar in material and style, but always differing in esthetic value in a direction agreed upon by art experts. Such pairs, of course, reduce the possible influence of subject matter on a child's choice. Confronted, for example, with two drawings of horses, the child could not say he chose one because he likes a horse better than a lion. Nor was subject matter the only potential influence whose strength was lessened in our study; there was some tendency, too, for a pair to comprise works resembling each

other somewhat in color, general shape, and other formal characteristics. By decreasing these various relatively simple contrasts, we hoped to discover whether the complex variables experts use in esthetic judgment also play a significant part in children's preferences. Despite our having selected stimuli in this way, we found that the simple contrasts between pictures continue to play an important role in the reasons children give for their choices.

Our focus on esthetic value also distinguishes this research from its predecessors in a second way--the choice of people to study. We have not interviewed a representative or random sample of children at each age. We have instead sampled at each age two groups of children whom we call high-scorers and low-scorers. The former are children whose preferences in our pairs agree most with esthetic judgments by experts; the latter are children whose preferences agree least with experts' judgments. Comparison between these two groups of children may help in understanding how development leads some children and not others to make an esthetic approach to art. (The only previous research at all similar in this respect is by Lark-Horovitz, 1937. She compared a special group of highly trained and presumably very talented children with relatively unselected ones. Selection was determined by the children's artistic productions, however, and we do not feel able to judge confidently how the two groups would differ as viewers of art.)

In the main presentation of our results, we consider variations in justifications for choice according to age and scoring group. We also have information on the way justifications vary with sex; since that variation is independent of the main developmental trends reviewed here, we discuss it only briefly at the end.

Subjects and Procedures

Subjects were 22 boys and 22 girls from grades 1 through 6, and 40 boys and 40 girls from grades 7 through 12, all evenly divided between high- and low-scorers. They were selected from several schools where we were surveying art preferences (Child, 1964).

In a New England city of about 150,000, we were studying preferences in four elementary schools, two in neighborhoods of relatively high social status and two in low. Only in the high-status schools did we find children who displayed a clear tendency to agree with expert judgment; for the interview study, therefore, we used only these two elementary schools. Secondary-school preferences were taken in two diversified suburban towns, when access to secondary-school students in the city proved unavailable. Since the distribution of preferences in these two towns was similar, interviews were conducted in only one of them.

For the survey of preferences, each child in the schools had indicated his personal choice in each of at least 130 pairs of slides made up as we have described. A total of 900 pairs were used. The children were not told how the pairs had been composed, and nothing was ever said to them, in survey or interview, about quality or value in art. Each child's 130 preferences were scored for amount of agreement with expert judgments. Classification of a child as high-scorer or low-scorer was in no way influenced by his performance in the interview; it was determined entirely by the preferences he had expressed in the group survey several weeks or even months before.

The names of the highest-scoring boys and girls in each grade, those we definitely wished to interview, were given to school authorities with the request that each student be matched for sex, grade and, as closely as

possible, general academic potential with one from a small list of low-scorers submitted at the same time. In the secondary schools, the three highest-scoring boys and the three highest-scoring girls in each grade were thus matched with low-scorers. A few of the students first picked and matched were lost to us because they moved away during the school year, refused consent for the interview, or had scheduling difficulties. Substitutions were made in the prescribed way. In each elementary school, the one highest-scoring boy and girl in each grade were matched with low-scorers, and no replacements were necessary. First-graders from only one elementary school were interviewed. One of the elementary schools included 7th and 8th grades; the eight interviews from these grades were analyzed with those of the junior-high students.

During the individual interviews, each child was shown a standard 30 pairs of slides, picked for this purpose from the 900 pairs used in the group-survey sessions, and was asked to state reasons for his preference in each pair. Responses were written down by the interviewer at the time, word for word so far as possible. Pairs included abstract paintings, portraits, landscapes, still-lives, drawings, sculpture, photographs of buildings, etc. In addition, secondary-school students were asked to fill out a questionnaire on personality characteristics (see Child & Iwao, 1968).

Analysis of the children's reasons was made independently by two judges, each working from an instruction manual prepared in advance by the authors, one of whom was also a judge and had acted as interviewer. Each child's statements about a pair were recorded on a separate card, and the 124 cards for one pair were analyzed in sequence. The judge who had been the interviewer sometimes remembered what child had made a given statement and the way it had been made. The second judge could be influenced only by the typewritten statement and his own inspection of the pair of slides to

which it referred.

Since our inquiry is concerned principally with how reasons vary from one child to another in relation to his age, scoring group, and sex, the most pertinent consideration about inter-judge agreement is the amount of agreement about how many of a child's 30 responses fall into a particular category. If a child tended to make statements that were ambiguous with respect to the category, one judge might place some in it and the other judge might place a slightly different set there. Thus there might be more agreement about the person than about an individual statement. It was agreement about the person that we decided to measure. Inter-judge correlation was generally satisfactory. For the 22 categories which averaged at least one mention per child, the inter-judge correlation (uncorrected) ranged from .38 to .97, with a median of .75. Some of our results, as will be seen, were obtained through pooling related categories which to some extent confronted the judges as alternatives; the reliability of the figures used for these results is probably substantially higher than it is for the individual categories.

In general, each judge's analysis led to the same findings. (To be certain, we worked up a preliminary form of results separately from the ratings of each judge.) We have, therefore, considered only the combined ratings of both judges. We began with a count of the number of times any judge considered a child to be making a response of a particular kind, and at the end of calculations we divided by two; thus the numbers reported here refer to average number of responses by the child, not number of recognitions by two judges.

The instruction manual prepared for the judges took into account some categories emerging from theoretical considerations and from similar

research previously published. Principally, however, it was developed by working through a sample of interviews and seeking to cover all features we could recognize as potentially frequent occurrences. Our present discussion is organized around six sets of general categories:

1. Reasons for liking vs. reasons for disliking, and total number of reasons given.
2. Reasons pertaining to color and light.
3. Reasons pertaining to inherent characteristics other than color and light (line, form, texture, etc.).
4. Reasons pertaining to content or subject matter.
5. Reasons pertaining to emotional meaning.
6. Miscellaneous reasons.

The grouping of specific categories into sets is inevitably somewhat arbitrary. Since, however, we have looked at results separately for each specific category, and have reported them wherever significant, the grouping into sets is primarily a convenience of presentation and does not determine conclusions.

A general problem we anticipated was that of inferring the sense or intent of an unclear comment. We attempted to deal with it by including this passage in the instruction manual:

The child's words should first be taken at face value as the conscious reasons for his preferences, and classified in the appropriate category. Some words frequently used are ambiguous; if reference to the slides being discussed indicates clear and indisputable reason for further, more specific classification, classify appropriately. Sometimes there seems to be reason to believe that words used by children are suggesting ideas too complicated for language available

to children. The principle for codifying them would be to codify first according to the words actually used, and then to codify in the classification for which there is ample reason for interpretation or inference.

A possible consequence is that one aspect of a statement might be counted twice; but we do not believe this was common.

When we present results on large sets of reasons given by children, we will do so for each of four age groups. Because, as will be shown, age has an irregular relation to total number of reasons, we will base that part of our presentation on relative rather than absolute frequency. That is, for each child we will begin by determining what proportion of his total reasons fall into the categories we are considering. Our tables will present the mean of this proportion for the children in each group.

We also present results for some of the single categories which make up the sets of reasons. Where a single category goes unused by many children, its frequency distribution is badly skewed and non-parametric statistical methods are needed. We have chosen chi-square and have decided to treat all variables as dichotomies (splitting always as nearly as possible at the median). Even though some of the specific categories occur rarely, they may serve to reveal differences between groups just as effectively as categories of much more frequent occurrence. In this analysis we put together as one body our two younger groups, and the two older groups as a second body, forming a contrast between elementary- and secondary-school pupils. There is no need to seek an adjustment for total number of reasons because in these two large age groupings the mean number of reasons given per child is almost identical. We analyzed 94 reasons (each consisting of a justification for liking and the corresponding justification for disliking--e. g., liking

because light or disliking because dark). Of these, 9 were referred to by 5 persons or fewer--too seldom for even a perfect correlation to be statistically significant.

Liking vs. Disliking, and Total Number of Reasons

Each reason a child gave was classified as a justification for liking one work more or the other less. Our subjects, we find, usually give affirmative justifications; they justify their choices by reasons for liking. Negative reasons, reasons for disliking, never reach in any age group as high as 20% of the total number of reasons. (This percentage may be seen in parentheses in Table 1.) An explanation may be that the children had been asked in the original survey, and were now again asked in the interview, to indicate which work they liked the better, thus suggesting statements about the better-liked rather than the less-liked.

If the general percentage of negative reasons is likely to have this procedural origin, variation of the percentage would be of greater interest. But we find no evidence that the percentage of negative reasons varies systematically with scoring group, and no conclusive evidence that it varies with age. As may be seen in Table 1, the percentage of negative reasons is lowest in the youngest group, and there may well be a consistent trend here; but the over-all age variation of this percentage is not quite significant at the 5% level ($F = 2.54$ with 3 and 108 df).

Results differing from ours are found in one previous study containing comparable information. In his study of French children, Francès shows a much higher general frequency of negative reasons (24% in his younger group, 36% in his older group, no more than 20% in any of our groups) and finds negative justifications to vary with age. The proportion of negative

Table 1

Total reasons: Mean number a child gives, and percent negative

		A g e G r o u p			
		Grades	Grades	Grades	Grades
		1-3	4-6	7-9	10-12
Scoring Group	High	53.7 (13.3%)	73.9 (19.2%)	65.4 (15.8%)	71.8 (18.4%)
	Low	58.4 (8.1%)	73.3 (13.4%)	60.3 (15.5%)	64.2 (16.4%)

reasons is decidedly higher in his subjects aged 13-16 than in those aged 6-12. The difference from our results might be connected with either cultural or procedural factors.

For total number of reasons, positive and negative together (appearing as the main set of entries in Table 1), we find variation with age, but not a regular progression. The children in the first three elementary grades give the fewest reasons, and the children in grades 4-6 give the most. Junior- and senior-high-school students fall between the two extremes. The variation with age is significant at the 1% level ($F = 4.81$ with 3 and 108 df). There is no significant difference between high- and low-scorers, nor is the apparent interaction between scoring group and age at all significant.

The change we note in total responsiveness, varying with age but not in a regular progression, is somewhat like that seen by Machotka. In the French children he interviewed about art preferences, he found a fairly regular increase with age in number of reasons given by his 6-to 12-year-old subjects. Above that age, Machotka interviewed only 18-year-olds, and the number of reasons they gave fell between the oldest and the youngest of the earlier age range. Francès' results, on the other hand, suggest number of responses steadily increases with age from 6 to 16 years. It may be important that the older children in Francès' study wrote their reasons. If there was no time restriction, they may have felt free to express all they felt and thought, and have given more reasons than some of our subjects who in giving their reasons orally may have imagined the interviewer to have only a limited time available for listening to them.

Color and Light

Various assertions that young children are especially interested in color and light might lead us to an expectation that our younger subjects would base their choice on color and light more frequently than would our older subjects. References to color and light are common in all groups of children, forming a little over 20% of all responses. But we do not find a consistent relation to age, and some specific color and light categories are used more by the older than by the younger children. There is, however, a very consistent relation to scoring status.

References to color and light were totaled separately from other formal properties of art. The reasons ranged widely: "It's colorful," "It has too great a variety of color," "I like the way the colors blend," "Blue is my favorite color," "It's lighter," "More pastel colors," "Too dark," "Good color scheme." The percent of references to all aspects of color and light is summarized, as a function of age and scoring group, in Table 2. No significant relation to age is present. But the very consistent relation to scoring status noted above is significant at the 1% level ($F = 4.36$, with 3 and 108 df). In every age group, color is mentioned less often by the high-scorers than by the low, though in the youngest group the difference is very slight.

We must now ask whether this finding is characteristic of all the separate categories of color and light references summarized in the over-all measure, or whether various specific categories have distinct tendencies that differ greatly from the total findings. Results for most specific categories cannot be confidently interpreted, since so few children use them. Generally speaking, the results for the specific categories do not seem inconsistent. That is, the low-scoring group tend to refer to some aspect of

Table 2

Reasons referring to color and light:
Mean percent of all reasons a child gives

		A g e G r o u p			
		Grades	Grades	Grades	Grades
		1-3	4-6	7-9	10-12
Scoring Group	High	22.0%	15.6%	20.1%	19.1%
	Low	22.1%	21.1%	26.5%	23.5%

The reasons for liking included in this table, and the percentage each one (together with a corresponding dislike) forms of the total 8038 reasons are, in descending order of frequency: Unparticularized reference to color, 6.6%. Colorfulness or color variation, 2.6%. Brightness (word "bright" used positively), 2.1%. Light-dark variation, 1.8%. Lightness, 1.6%. Blending of colors, 1.6%. Darkness, 1.1%. Cool hues, 1.0%. High saturation, 0.7%. Lighting, 0.7%. Warm hues, 0.5%. Specific hues not clearly cool or warm, 0.4%. Lack of brightness, 0.4%. Color uniformity, 0.2%. Light-dark uniformity, 0.1%. Low saturation, 0.1%. Clashing of colors, 0.1%.

color or light more often than the high-scoring group, and there is usually no consistent variation among age groups. To the first generalization there is only one significant exception, which we will consider later. To the second generalization--the lack of an age effect--there are three exceptions. These are not in a uniform direction and work against one another in contributing to the summary figures of Table 2.

One exception to the generalization on age effect consists of an unparticularized, unqualified reference to color as a reason for liking, a reference that does not specify any aspect of color. These references took such forms: "I like the colors in it," "Because of the colors," "I like the colors better," or, laconically, "Colors." These unmodified references to color increase regularly with age; they are given at least 4 times by 65% of the older children, but by only 39% of the younger. (Here and later, when we cite a given number of references, we are simply citing the number nearest to the median, selected for the most adequate statistical test.) The variation with age is highly significant, with $\chi^2 = 7.41$ (χ^2 , for our use in this report, is significant at the 5% level if it exceeds 3.84, and at the 1% level if it exceeds 6.64). Is this fact contrary to what we might expect? Before answering, let us consider an opposite finding.

Here we look at the specific category of liking pictures characterized by colorfulness and variation in color. In this category we entered such phrases as "Colorful," "More colors to it," "Many colors," "Has more color." (These phrases doubtless at times mean high saturation, but only when sure of this meaning did we use the category of liking because of saturation, and the results for that category itself are not significant.) The greatest variation is with scoring group; colorfulness and color variation are given as reasons more than once by 59% of the low-scorers, but by

only 27% of the high-scorers ($\chi^2 = 11.73$). In addition, however, we note a sizable age effect, with colorfulness and color variation frequently cited by 57% of the younger and by only 36% of the older children ($\chi^2 = 4.10$). Our results thus confirm the view that younger children like the vivid assaultive character of strongly varied color. The tendency for unparticularized references to color to be given more by older children, then, obscures or cancels out in the summary of Table 2 the fact that younger children like strong variations of color.

The third significant finding on age in relation to specific categories further illustrates the complexity of the general category of color and light. Preference for a work because it is less bright than its mate (for example: "I don't like bright and gaudy colors," "It's not so bright.") is stated by 27 children, all but one in secondary school. This highly significant difference ($\chi^2 = 13.50$) suggests that younger children may like color and light more; the evidence, paradoxically, is provided by older children's explicit rejection rather than younger children's explicit acceptance.

Our results may help to explain why previous studies have differed from one another in their findings on color. Machotka finds decreasing reference to color with increasing age. Francès finds little change--some increase in absolute numbers, a slight decrease in percentage of all reasons. Lark-Horovitz reports little change up to the age of 10 or 12, a slight decrease thereafter. In the differing results we obtain for two particular categories of color lies evidence that various kinds of reference to color may differ widely in their relation to cognitive development of the child. Liking "colorfulness" may be a fairly simple attraction to what is most immediately and primitively conspicuous in a work; liking the "color" of a

work may be an inadequate verbal representation of a complex appreciation. Of the various inquiries--those by Machotka, Francès, and others--some may, because of their stimuli or other aspects of procedure, have more strongly called attention to colorfulness or its absence and thus led younger children especially to mention color. Others may have used stimuli which especially prompted older children to comment on the role of color in the total work. Discrepant age relations might thus have resulted from procedural differences.

One other specific category in the over-all classification of color and light has a significant finding. It is the one referred to earlier as providing an exception to the generalization that low-scorers mention color and light more often than do high-scorers. High-scorers, we find, are more likely than low- (66% as against 45%, $\chi^2 = 4.71$) to give as a reason for choice the blending of colors in a painting. This kind of color reference is concerned with complex relations among elements of a picture; in this respect it resembles references to other characteristics to be discussed in the following section, also occurring more frequently in older children.

Inherent Characteristics other than Color and Light

Under this not-too-satisfactory label we have grouped together the most purely formal characteristics (such as symmetry vs. asymmetry, depth, shininess, clarity) and other characteristics which may be influenced by subject matter but are largely of formal origin (such as softness vs. hardness, strength vs. delicacy, activity vs. quiescence). This offers a convenient comparison with the general category of color and light, and with the other general categories we consider later. Altogether these reasons comprise a little more than 45% of the total. As may be seen in Table 3, this over-all group shows most conspicuously a highly significant variation

Table 3

Reasons citing inherent characteristics other than color:

Mean percent of all reasons a child gives

		A g e G r o u p			
		Grades	Grades	Grades	Grades
		1-3	4-6	7-9	10-12
Scoring Group	High	30.2%	43.1%	47.9%	50.2%
	Low	35.8%	50.3%	43.7%	49.4%

The reasons for liking included in this table, and the percentage each one (together with a corresponding dislike) forms of the total 8,038 reasons, are: Photographic realism, 6.4%. Detail, 3.7%. Composition, 3.6%. Background, 3.3%. "More in it," 2.4%. Shape, 2.3%. Clarity, 2.3%. Simplicity, 2.1%. Activity, 1.7%. "Less in it," 1.5%. Style, 1.5%. Curvature or gracefulness, 1.4%. Shading, 1.3%. Pose, 1.0%. Depth, 1.0%. Concentration, 0.9%. Lines, 0.9%. Order, 0.8%. Unity, 0.8%. Softness, 0.7%. Strength, 0.7%. Verticality, 0.6%. Complexity, 0.6%. Diffusion, 0.6%. Completeness, 0.5%. Symmetry, 0.4%. Thinness, 0.4%. Absence of detail, 0.4%. Quietness, 0.3%. Straightness, 0.2%. Proportion, 0.2%. Roughness, 0.2%. Smoothness, 0.2%. Texture (unspecified), 0.2%. Delicacy, 0.2%. Horizontality, 0.1%. Asymmetry, 0.1%. Disorder, 0.1%. Shininess, 0.1%. Dullness, 0.1%. Incompleteness, 0.1%. Thickness, <0.1%. Flatness, <0.1%. Obscurity, <0.1%. Absence of unity, <0.1%. Hardness, <0.1%.

with age ($F = 17.70$). For the low-scorers, the variation with age takes the same form as do the total results presented in Table 1, even though the entries here in Table 3 consist of percentages of those in Table 1. The tendency toward high responsiveness we have already noted in children of Grades 4-6 is reflected especially strongly, then, in the low-scorers' use of reasons pertaining to these inherent characteristics. For the high-scorers, on the other hand, the change with age appears as a regular progression, a steady increase with age in the relative frequency of these types of reasons. This interaction between age and scoring group does not quite reach the 5% level of statistical significance (F is only 2.33 instead of the approximately 2.70 required). The simple relation to scoring group does not approach significance (F is only 1.27).

This summary, however, embraces a large number of specific categories, and when we turn to considering them separately a somewhat different impression is gained. The tendency for increase with age occurs almost uniformly, and is significant for 11 of the separate categories. There are several insignificant reversals, and a single significant one which we will consider shortly. But almost as many of the relationships are significant for scoring group--10 of them! These, too, are mostly in a consistent direction, indicating that high-scorers cite inherent characteristics more often than do low-scorers. Here we find two significant reversals as well as several that are not significant. The reversals are a useful place to begin, as a precaution against over-generalizing the main findings.

The significant reversal in relation to age appears in the quantitative criterion--liking a picture because "There's a lot in it," "Not just one horse, a lot of horses," "The artist shows more," etc. This criterion shows a relation to age exactly opposite to the main effect. Of the

secondary-school children, only 31% offer such reasons more than once, whereas of the elementary-school pupils 66% do ($\chi^2 = 11.53$). The inherent characteristics of art influencing evaluations by young children, then, may be simply those to which the evaluative scales of everyday life can be most readily and obviously generalized. We would venture to guess that if our stimuli had not been approximately standardized in size through being prepared as slides, the larger in a pair might have been chosen on similar grounds.

The two inherent characteristics atypical in being cited more frequently by low-scorers than by high are a liking for clarity ("I like it because it's easy to see," "It stands out more," "You can see it better.") and a liking for sharp, clearly defined detail ("The lines are more distinct, not blurred," "It doesn't look so smudged," "The other's hazy."). Liking clarity or disliking lack of it was stated once or more by 61% of the low-scorers and by only 39% of the high-scorers ($\chi^2 = 5.45$); liking sharp detail or disliking its opposite was stated at least twice by 55% of the low-scorers and by only 35% of the high ($\chi^2 = 3.94$).

High-scorers, it would appear, tend not to admire what is very detailed and perfectly clear. Some of the inherent characteristics they name more often confirm this point. One is the opposite of liking sharp detail: preferring a picture because it is not so detailed as its mate. This reason is mentioned by 39% of the high-scorers and by only 19% of the low ($\chi^2 = 4.74$). Consistent with this is the finding for disorder as a basis for choice (e.g., "I like it because of the cluttered effect," "The other looks too neat," "I don't know why, it's just nice--jumbled."). Disorder is mentioned by only 15 of our subjects altogether, but 12 of them are high-scorers, and the difference is significant ($\chi^2 = 4.85$). Probably related

to this group of reasons is preference for the more active picture in a pair ("It has so much movement in it," "It has more prancing and running in it," "It gives more a variety of action," "Action and movement."). Activity is cited once or more by 66% of the high-scorers and by only 39% of the low. Possibly related, too, is liking roughness (cited by 22% of high- and 5% of low-scorers, $\chi^2 = 6.82$) and liking texture (18% vs. 5%, $\chi^2 = 3.95$) if we assume that a roughness of texture is usually what is implied.

The high-scorer thus seems to prefer haziness and lack of detail, coarseness of texture, and a sense of activity. Into this fairly consistent view can perhaps be fitted two other findings. One is that high-scorers are more inclined than low-scorers to cite greater visual depth as a reason (mentioned by 58% as against 39%, $\chi^2 = 3.91$). We might regard successful representation of depth on a flat surface as an instance of supplying greater detail and clarity, and consider this a paradoxical finding. On the other hand, convincing experience of depth when looking at a two-dimensional surface may be thought of as suggesting activity and disorder, and thus the results may be interpreted as consistent with each other. A similar finding is that high-scorers are more likely to prefer the more curved or graceful art (63% of the high- and 37% of the low-scorers offer such a reason once or more, $\chi^2 = 7.26$). This might be thought of as a contrast to liking rough texture, but it may also be regarded as related to liking activity.

One finding on scoring group, though, cannot be so readily reconciled with the rest. High-scorers are more likely than low-scorers to give simplicity as a reason for liking (once or more by 55% of high and by 35% of low, $\chi^2 = 3.94$). Simplicity seems hardly the message of the previous cluster of reasons, though to be sure complexity is not really implied

either, except by the one category of disorder. Putting together these several findings suggests, as is probably true, that high-scorers have different reasons for liking different pictures, that they are in this sense more discriminating or differentiated than low-scorers in their response to art. Our present data provide no adequate general test of this hypothesis. One of our high-scoring subjects almost stated it herself, though. She had given simplicity as a reason for choosing one vase over another; right afterward, justifying her choice of one rug over another, she commented, "It's more complex; isn't it funny how I change?"

We have now presented 10 reasons significantly correlated, in one direction or the other, with scoring group. Only three of these--roughness, unspecified texture, and lack of detail--are significantly correlated with age. They are correlated with age in a corresponding direction; that is, the reason associated with high score is also associated with greater age. The temptation becomes great to conclude that younger children and low-scorers completely parallel each other, and that older children and high-scorers do likewise, in esthetic response. But it is useful to consider some quantitative data that are not themselves statistically significant. Detail, for instance--a reason for liking which characterizes low-scorers--shows no parallel tendency to be given by younger children; it is actually given slightly more frequently in our secondary-school than in our elementary-school subjects (48% against 41%). A similar discrepancy occurs with activity, a reason given significantly more often by high-scorers than by low; it tends to be given more often by younger children (57%) than by older (50%), though again this difference is not significant. However, the five other specific reasons we have already considered do show a trend with age--insignificant statistically--that corresponds in direction to the

significant difference between scoring groups.

Seven reasons in this over-all classification remain to be discussed. They show a significant relation only to age group; but, though not statistically significant, the relation to scoring group follows in the corresponding direction in all save one instance. These remaining reasons, all given significantly more often by our subjects in secondary school than by those in elementary school, are rather difficult to characterize as a group. One of them, liking what is smooth, indicates perhaps that the adolescent's interest in texture, expressed by high-scorers in liking roughness, is not confined to that end of the dimension. Smoothness is mentioned as a reason for liking by 20 of our subjects, only 2 of them in the elementary schools ($\chi^2 = 5.50$). Four reasons form a set having in common, perhaps, reference to rather recondite concepts not likely to be expressed until a child has acquired a considerable vocabulary--and perhaps not until he has heard them applied to the discussion of art. One of the four is symmetry, given as a reason by 22 subjects of whom one alone was below the seventh grade ($\chi^2 = 9.60$). Another is unity, mentioned by 46% of the older group and by 16% of the younger ($\chi^2 = 10.13$). The word unity was itself often a part of the statement, but also used were such phrases as, "Grouped so I see it as a whole, not as individual parts," "All put into one," "Things look more closely related." A third concept or category here--a frequent one--is composition or arrangement, referred to at least twice by 58% of the older children and by only 30% of the younger ($\chi^2 = 8.48$). Examples are, "Arranged better," "Composition better," "The way the table is, and the way they put the things on it," "You can tell there's a body to it, but the light head makes you go right to the important part of the body." The fourth category is preference for the more concentrated, centered, or

focussed (e. g., "Not so spread out," "Things are closer together," "The fruits are arranged in such a way that you don't have to look all over the picture to see them."). Reasons so classified were given by 48% of the older subjects and by 25% of the younger ($\chi^2 = 9.60$).

Two categories have a possible linkage that may be worth mentioning though it cannot be adequately tested. One is an impression of softness as a reason for liking, mentioned by 36% of the older group and by 16% of the younger ($\chi^2 = 4.76$); the other is a feeling of strength, mentioned by 53% of the older and by 32% of the younger subjects ($\chi^2 = 4.10$). These aspects of the connotative meaning of art may well serve as sexual symbols--softness as a symbol of femininity, strength as a symbol of masculinity. If they do indeed represent sexuality--whether in a narrow sense of genitality or in a broader sense of socially defined separation of role--it seems probable that they would influence adolescents' preferences more than those of younger children. This hypothesis leads us to inquire about other inherent characteristics that might also serve in sexual symbolism. Two have already been named in another connection, both having been reported to occur more in our older subjects: curvature or gracefulness (though the age difference here is not significant), and roughness--possible symbols, respectively, of femininity and masculinity. Among the categories failing to show any significant relation to age there are, moreover, only four which seem to qualify as potential sexual symbols: delicacy, for possible feminine meaning; verticality, straightness, and hardness for possible masculine meaning. Of these, all but verticality are rare enough so the lack of significant findings is hardly disturbing. The evidence seems to encourage further exploration of symbolic sexual meaning as a possible factor in age changes.

Our general category of inherent characteristics other than color

and light seems to overlap greatly with Francès' category of "technique." Comparison of results may be warranted: Francès' tables show that references to technique as a reason for liking are more frequent--though not by a large margin--in his older group, those 13 years or above. Whether this is a significant finding we cannot tell, as Francès does not discuss age variation in this justification for choice; the direction of the difference is, however, confirmed by our own statistically significant findings. Nor do we know whether the curvilinear relation apparent in Table 3 would also be found in Francès' data, since his subjects were divided into only two age groups.

Content or Subject Matter

Deciding whether the child gives subject matter as a reason for preference is not always easy, for subject matter is often mentioned in stating other reasons. Choosing one still-life to another because "I like carnations better than roses" clearly gives subject matter as the reason. Justifying the same choice by saying, "I like the picture of carnations better because they're painted a deep red color that I just love," seems to refer to subject matter just to express a color preference. If the child says only, "Ooh, those carnations are so red," the analyst has difficulty deciding whether the subject matter should be considered a reason for choice. An initial attempt to include references to subject matter in the general scheme of analysis was frustrating both for this reason and because it was distracting to judge separately what was being said simultaneously about subject matter and the form in which it was represented. We decided to isolate and quantify the task of judging whether subject matter was cited as a reason. After the general analysis was completed, each judge independently went back through all the statements to rate each on the

following scale: justifying choice by reference to subject matter (2), not doing so (0), or being ambiguous (1). This measure for individual children (summed, that is, over the 30 pairs on which they commented) showed excellent inter-judge agreement, with a correlation of .94. Since it is a separate measure cross-cutting the rest of our analysis of reasons and only indirectly related to total number of reasons, there seems no need to seek a suitable way of relating it to total responsiveness, and we will deal with it only absolutely.

Use of subject matter to explain preference is most common in young children and decreases greatly with age, but shows no consistent relation to scoring group, as may be seen in Table 4. A big drop in references to subject matter comes between elementary and secondary school. The age variation is highly significant ($F = 19.4$); it is especially impressive when we recall that our subjects were always justifying a choice between two pictures similar in subject matter.

The age variation we find echoes that obtained long ago by Lark-Horovitz who, on asking children to explain their choice of favorite picture out of a number shown them, found a marked decline in subject-matter reference even between the ages of 6 and 10. Our results confirm, too, those of Frances. He has a category for subject matter, and indicates that it is used a great deal more by his children aged 6-12 than by those 13-16. Machotka offers a more differentiated treatment of subject-matter reasons. Those references that seem to him to express an empathetic identification with the subject matter--as though the child looking at the picture did not sharply distinguish himself from it--Machotka finds more frequent in his youngest children (except at age 6, which may be a sampling error) and especially scarce in his one late-adolescent group, the 18-year-olds. The

Table 4

Summed ratings on reference to subject matter
in justifying 30 choices: Mean per child

		A g e G r o u p			
		Grades	Grades	Grades	Grades
		1-3	4-6	7-9	10-12
Scoring Group	High	39.0	38.9	24.0	25.0
	Low	39.0	35.8	22.5	25.6

special decline of the simple empathetic reason is viewed by Machotka as consistent with the decline of egocentricity in general cognitive development. Its rapid decline around the age of 12, he interprets in accordance with Piaget's view that this is the age when "the final loosening of thought from its dependence on concrete data takes place." Machotka reports that miscellaneous references to subject matter also decline, though more gradually, with increasing age. He finds, however, that references to the affective tone of the picture, clearly localized in the picture rather than in the child, show an opposite effect, increasing rather than decreasing with age. Results we will report in the next section seem consistent with this last finding.

Emotional Meaning

Reasons for preference that refer to emotional meaning, we felt, could be divided into two groups corresponding fairly well to the common terms, sentimental and emotional. These terms did not appear in the manual as categories, but we have subsequently grouped under each a number of pertinent specific categories recognized in the manual. Each group comprised about 10% of the total reasons analyzed.

The two groups seemed very different to us, and so they were in outcome. Reasons we classified as sentimental are related principally to age, as may be seen in Table 5. They are most often offered by the youngest children, and decrease in use up to secondary-school age, when further change seems to be uncertain. The age change is highly significant ($F = 11.49$, with 3 and 108 df); no other sources of variation even approach significance. The over-all results for emotional reasons are quite different, as is shown in Table 6. The overwhelming variation here is between scoring

Table 5

Reasons classified as sentimental: Mean percent of all reasons a child gives

		A g e G r o u p			
Scoring Group		Grades 1-3	Grades 4-6	Grades 7-9	Grades 10-12
High		15.2%	9.5%	6.8%	5.6%
Low		16.8%	9.2%	8.2%	8.4%

The reasons for liking included in this table, and the percentage each one (together with a corresponding dislike) forms of the total 8,038 reasons, are: Pleasantness, 4.0%. Prettiness, 3.3%. Familiarity, 0.8%. Normality, 0.3%. Emotional restraint, <0.1%.

Table 6

Reasons classified as emotional: Mean percent of all reasons a child gives

		A g e G r o u p			
Scoring Group		Grades 1-3	Grades 4-6	Grades 7-9	Grades 10-12
High		8.5%	11.0%	14.1%	13.6%
Low		4.8%	5.4%	7.6%	7.3%

The reasons for liking included in this table, and the percentage each one (together with a corresponding dislike) forms of the total 8,038 reasons, are: Presence of feeling, 3.6%. Trueness-to-life, 2.2%. Interestingness, 1.8%. Unfamiliarity, 1.4%. Stylization, 0.5%. Unpleasantness (tragedy, suffering), 0.4%. Abnormality (distortion, eeriness), 0.2%. Ugliness, <0.1%.

groups ($F = 22.57$); very consistently, at all ages, the high-scorers give responses bearing on emotion about twice as often as do the low-scorers. The simple variation with age is also statistically significant, but just at the 5% level. The change with age is directly opposite to that for sentimental justifications; emotional justifications appear least frequently in the youngest children. The effect of age and scoring status on emotional and sentimental reasons produces large differences in the extreme groups: the youngest low-scorers give sentimental reasons $3 \frac{1}{2}$ times as often as emotional; the oldest high-scorers give emotional reasons about $2 \frac{1}{2}$ times as often as sentimental.

Sentimental reasons include a liking for what is pleasant, pretty, or unthreatening to emotional equilibrium. Major contributions to the total result are made by two specific categories, each showing results consistent with the total finding. One of these categories is liking what is pretty and disliking what is ugly--"The children are cuter," "The woman is more attractive," "The colors are prettier." Such reasons were given more than once by 68% of the elementary-school children and by only 39% of those in secondary school ($\chi^2 = 8.14$). The other specific category with significant outcome was that of liking the familiar and disliking the strange--"Looks like a regular person you see anywhere," "You see the country like it really would be, not imagined," "Too grotesque." These reasons occurred in 64% of the younger children and in 38% of the older ($\chi^2 = 6.77$). Another category adding substantially to the total for sentimentality, although the results for it alone are not statistically significant, is liking the pleasant or happy and disliking the sad or unpleasant--"She looks very happy, as if she didn't have a care in the world," "The fruit looks as if there's going to be a party," "The children look like they're enjoying themselves,"

"In the other the horses look as if they're being trained, not free to enjoy themselves romping in the fields."

The general category of emotional reasons contains all references to liking seeming depth or genuineness of feeling--the presence of strong emotional evocation--and to disliking shallowness or concealment of feeling. Here three specific categories, each with statistically significant results, contribute most to the total findings. Preferring the work with more feeling has the largest relation to scoring group. Typical phrases are, "The one on the right has more feeling. You could take a photo of what's shown on the left and get the same thing," "There's more to it--it's not as empty as the other," "You can get an emotion from it," "More of an expression on her face. The other just looks happy." These reasons are given at least twice by only 23% of the low-scorers but by 71% of the high-scorers ($\chi^2 = 27.24$). They are also more frequent in the older children (54% against 34%, $\chi^2 = 4.12$).

The second specific category consists of favoring the work that seems more true to life. Included were such comments as: "This has more feeling of a garden and actual flowers," "The face has a human expression. Very true to life," "The kids are real kids, not just pictures of children." This category was found more than once in only 39% of the low-scorers but in 69% of the high-scorers ($\chi^2 = 10.53$). It was also more frequent in the older children than in the younger (61% vs. 41%, $\chi^2 = 4.30$).

While both these specific categories show their strongest relation to scoring group, they do also, as we have indicated, show a relation to age. A third category shows a relation only to scoring group. This is the justification of choice by reference to stylization, i. e., some expressive distortion of reality, and the rejection of photograph-like copy of surface

appearance. For example, "It gives the artist's impression; it has more feeling." "This looks like something you feel about a place; the other looks too realistic, like with a camera--it doesn't mean much," "It looks like a photo rather than a drawing." Such reasons are mentioned by 44% of the high-scorers and by only 15% of the low ($\chi^2 = 11.31$). Relation to scoring group appears equally strong in younger and older children.

These results cannot be directly compared with those of Machotka and Francès, since neither makes an explicit contrast between emotional and sentimental. Machotka's distinction, however, between two types of content reference, already mentioned in the preceding section, seems likely to be relevant. In many instances, the empathetic identification may well be recognized through statements we would place in one of our specific categories classified as sentimental. Justification by affective tone seems likely, on the other hand, to consist of reasons we would usually put under emotional. If this view is correct, our results on age changes agree with Machotka's. Francès has a category of expressivity that seems to have large overlap with our category of emotional reasons. Here, too, is agreement, for Francès finds such reasons to be given much more frequently by children 13-16 than by children 6-12.

Miscellaneous Reasons

Some of the reasons we analyzed, and found to occur often enough to justify statistical treatment, do not seem to belong in any of the general categories around which we have organized our presentation. One is liking the old and disliking the new. This does not refer to youth and age in the persons portrayed (reasons which were very rare) but to the past vs. the present, in either the subject matter or anything else about the picture that suggests this dimension to the child (e.g., "It's a more old-fashioned

home," "Early times," "Typical town of years ago," "Doesn't look as new-fashioned."). Liking oldness was found more often in high-scorers than in low (48% vs. 24%, $\chi^2 = 6.84$). It was also found more often in younger children than in older (57% against 25%, $\chi^2 = 11.09$). This is an unusual combination. Ordinarily, as we have said earlier, what tends to characterize the high-scorers also tends to characterize the older children as a whole; here, the reverse is true. Liking the past or oldness is especially typical, then, of the high-scoring elementary-school children; 77% of them give such reasons, whereas only 18% of the low-scoring secondary-school children do.

Also of interest, in connection with the criterion of oldness, is the fact that the relation of this reason to scoring group is much closer in the younger children than in the older. When we look at all our variables, we find a decided tendency for the opposite to be true. Here, then, is a criterion which appears especially relevant to art appreciation at the elementary-school age, whereas other criteria are more relevant at a later age.

Another standard perhaps pertinent to art appreciation only at younger ages is suggestiveness or evocativeness. Four reasons for preferring a male portrait by Hals to one by Rembrandt provide apt illustration: "He looks like a Pilgrim," "It reminds you of the old days--Pilgrims," "The one on the right is a Pilgrim; he has a nice hat and face, a nice-looking coat; he's holding a child by the hand," "Because he's like a fighter--a famous sword-fighter on a horse. He's wearing a black hat so the enemy won't recognize him." In the last two quotations, the child is actually including, as though describing the picture, elements present only in the story it has suggested to him. This category is used two or more times by

82% of the elementary-school children and by only 35% of the secondary-school children ($\chi^2 = 23.10$). In the secondary school these reasons appear with equal frequency in high- and low-scorers. In the elementary school the trend (not significant) is for them to be given more by high-scorers than by low. This fact is reminiscent of what we have described for oldness, suggesting that esthetic appreciation, despite its having some constant characteristics at different ages, may also vary decidedly with age.

One of the miscellaneous criteria provides the only instance, among all our variables, of strong evidence that a reason may actually have a reverse relation to scoring group at different levels of maturity. It is liking what is better suited to use, or disliking what is not so well suited to use. This criterion was most often elicited by a pair of chairs, though it also appeared in comments on a pair of bowls. Among elementary-school children, it seemed especially likely to be stated by high-scorers, though the relation did not quite reach statistical significance ($\chi^2 = 3.28$). Among secondary-school children, there was a significant tendency for low-scorers to use it more often ($\chi^2 = 4.14$).

Two additional miscellaneous reasons, correlated positively with score, pertain to knowledge about the picture or its origins. One is preferring a work believed to be valuable (e. g., "It looks as if it belongs in a museum," "You couldn't buy that in a discount house."). Such reasons were given by only ten of our subjects, all but one of whom were high-scorers (mention by 2% vs. 14%, $\chi^2 = 5.33$). The other is preferring work recognized--rightly or wrongly--as being by a famous artist (e. g., "Oh, that's by Cézanne!" "It looks famous," "Simply because I think I saw it before someplace," "It was done by a famous artist--I have it in a book."). Such a reason was given by 18% of the low-scorers and by 40% of the high

$(\chi^2 = 6.61)$.

More germane to some of the reasons we have reported in earlier sections is another miscellaneous one: liking a work because it offers a challenge to understanding. For example, "The more puzzled expression on his face makes you wonder more than the one on the right," "You get to wondering where he is," "Because of the men--it's intriguing, you don't really know what they are," "The mood of the color--more intriguing than the one on the left, which is only gay. You know what is happening on the left. The one on the right makes you wonder what they're thinking about. I like a painter that makes me wonder." This category was used by only 12 of our subjects; all but one of them were high-scorers, and this difference between scoring groups is highly significant ($\chi^2 = 7.47$).

A reason opposite to this--liking the easy-to-understand work--is, on the other hand, related significantly to age: 23% of the elementary-school children give such a reason, whereas 65% of the secondary-school children do ($\chi^2 = 18.64$). It might be expected that this reason would show a negative relation to scoring group, but this expectation is not fulfilled. The relation to score, though not at all significant, is in the positive direction, that is, there is a trend for this reason to be used by more high-scorers.

Sex Differences

The most striking finding about sexual differences is that they are very few. The age and scoring-group differences we have described for broad groups of reasons seem generally to be equally true for both sexes; that is, we do not find significant interactions between sex and either of these variables. For one of the broad categories of reason, however, sex does show a large effect independently of effects of age or scoring group.

This one substantial finding is that boys give a higher percentage of negative reasons than do girls. The means appear in Table 7. The sex difference is significant at the 1% level ($F = 7.12$). It appears to diminish with increasing age, but the interaction does not even approach statistical significance. Boys' greater use of negative reasons also creates a significant difference between the sexes in total number of reasons given, a difference attributable entirely to the negative reasons since positive ones are given with almost exactly the same frequency by the two sexes. This one large difference between the sexes seems to be an instance of the commonly observed greater aggressiveness and independence of boys. Knowledge of general sex differences might have led us also to expect some other large differences in reasons for choice. These do not appear. For instance, we might well have expected girls to give more readily both sentimental and emotional justifications; in fact, there is no sex difference in either of these general categories.

Turning to the separate categories, we find no clear evidence of sex differences. As indicated earlier, we applied the chi-square test of significance to 85 separate reasons for choice occurring frequently enough so that significant results were a possibility. Of these 85, just 4 showed a significant sex difference by the 5% criterion, and none at all reached significance at the 1% level. The direction of three of these differences seems reasonable: girls more frequently give brightness and softness as reasons, and boys more often say they like a relative uniformity of color. But the fourth difference, that boys are more likely to cite normality as a reason for liking, is counter to the expectation that boys would be more independent. So the character of these four findings hardly gives us any more confidence than does their fewness in believing that they represent

Table 7

Negative reasons: Mean percent of all reasons a child gives

		A g e G r o u p			
		Grades	Grades	Grades	Grades
		1-3	4-6	7-9	10-12
Sex	Boys	13.7%	18.1%	17.8%	17.5%
	Girls	5.9%	13.9%	12.0%	15.1%

any genuine tendency at all.

In contrast to sex, our main concerns in this paper--the relation of age and scoring group to reasons given for choice--do show numerous and large differences, as already indicated. Scoring group shows a significant relation to 19 out of the 85 reasons mentioned above, and with 9 of these the significance is at the 1% level. Age shows a significant relation with 21 reasons, and at the 1% level with 14 of them. Sex, then, is clearly not an important influence on the reasons a child gives for his art preferences; of greater weight are his general stage of development and his approach to art defined by his sharing or not sharing the preferences of experts.

Summary and Discussion

The single sex difference of which our results make us confident offers little for discussion. In a time when psychology is discovering unexpected and mainly puzzling differences between the sexes, it is refreshing to find an area of behavior where the sexes seem more alike than we would have supposed. And it is appropriate that art, as one of the humanities, should be an area where both sexes share a common humanity. If future studies of response to art continue to confirm the near-absence of sex differences, this will be an important point to take account of in a general psychological theory of art. At present we can best discuss the large differences we do find among age groups and between high-scorers and low-scorers.

We have studied children in each grade from first through twelfth (corresponding to typical ages from six through seventeen). In each grade we have taken, not a representative sample, but two extremes: High-Scorers, who tend to prefer art that experts consider esthetically better, and Low-Scorers, who tend to have contrary preferences. We find marked differences

between reasons offered by older and younger children, and we find marked differences between reasons offered by High-Scorers and Low-Scorers. A generalization can be made about the relation between these two effects, but it should be accepted only with reservations. That is, the two sets of differences are to some extent similar; reasons given more often by older children show some tendency (though usually not significant) to be given by high-scorers. And reasons given by high-scorers show some tendency (again usually not significant) to be given more often by older children. Significant variation with both age and scoring group is shown by only one overall category of reason (emotional) and only 8 single categories; in all but one of these instances the two variations go in the same direction in the way we have just described.

There is a general tendency for each effect to be present regardless of the other variable: Differences between low- and high-scorers tend to be the same in elementary school and secondary school; and differences between the two age levels tend to be the same in low-scorers as they are in high-scorers. The results for subdivided portions of our data would usually not be significant statistically. Generally, however, subdivided portions show differences in the same direction, and often of similar magnitude; there appear to be few significant interactions between the age and score dimensions. In the large categories to which we have applied analysis of variance, there are no significant interactions. While we have not applied any comparable test of interaction to the specific categories, there are few instances (we have described them all) where genuine interaction seems likely on our evidence.

Yet the two variables--age and scoring group--seem clearly to be different in their implications for children's reasons for choice. The

difference appears when we try to characterize the set of reasons showing significant and especially close dependence upon scoring group, and to compare this with the set of reasons showing significant and especially close dependence upon age. With very good though not perfect consistency, the two sets form distinct patterns.

We will begin with age. Younger children are especially likely to mention subject matter; the decline with age seems a likely consequence of increasing capacity for objectivity. Younger children are also prone to mention the evocativeness of art and to present a story or dramatic setting it suggests; here too the decline may be viewed as a symptom of increasing objectivity. As Machotka says in dealing with different kinds of subject-matter references, the young child responds with empathetic identification, while the older child can consider more objectively what is presented. The specific reasons favored by young children among the ones we have classified as referring to inherent characteristics are colorfulness and the quantitative criterion of presence of many things. These are both simple bases for evaluation: the conspicuously stimulating, the conspicuously ample. With greater age, children give more differentiated responses: statements about color which, though often vague, imply attention to complex relations, and discriminating statements about some of the other inherent characteristics. The older children often give reasons pertaining to much more complex features of art than the quantitative criterion; yet the way they allude to these more complex features indicates a continuing concern with simplicity, as in liking symmetry, unity, composition, and concentration. We can give similar meaning to the fact that ease of understanding--classified under miscellaneous reasons--was cited by older rather than younger children. This again implies recognition of complexity--here mystery and

intrigue--yet basically indicates a search or desire for simplicity. The older children also tend to offer another set of reasons, liking strength, softness, smoothness, roughness, and texture generally, which again imply a differentiated awareness of the objective characteristics of the art. As we have indicated earlier, these reasons may also suggest interest in characteristics symbolizing the division of humanity into male and female.

Younger children tend to give what we have called sentimental reasons for choice--liking the pretty, familiar, and pleasant, for example. They may be seen as a part of the relative simplicity of the younger children in their interaction with art; accentuating the positive may serve to ward off the danger of emotion in the simplest possible way. This tendency declines with age. To some extent, it is replaced by definite interest in emotion; but this is more closely associated with scoring group than with age, and the principal correlate of age in the emotional sphere is the decline, with years, in sentimentality.

In what they say about color and light, the scoring groups seem to differ in much the way we have described the age groups as differing. The low-scorers are more inclined to refer to simple colorfulness, the high-scorers to something more complex (in this instance, the blending of colors). For other inherent characteristics, however, the pattern of distinction between scoring groups is very different from that between age groups, though some of the correlations are similar. The low-scorers say they like clarity and detail; most of the high-scorers' reasons in this category indicate a liking for lack of detail and order. This pattern--low-scorers preferring the clear, direct, and simple, high-scorers liking haziness and disorder--is repeated and strengthened by the high-scorers' expressions of liking activity, roughness, texture, depth, and challenges to understanding.

The pattern is a familiar one in recent research on personality. It echoes Barron's portrayal (1958) of characteristics of more creative or original persons when compared with less creative ones. Child (1965) has shown that a similar pattern distinguishes American college students who evince an esthetic approach to art (by the same criterion we use here) from college students who lack such an approach. Child and Iwao (1968) show a similar distinction among college students in Japan, and report comparable findings from a personality questionnaire administered to the secondary-school high- and low-scorers whose interviews we have considered in this paper. The differences indicated in their questionnaires, we find, are indeed displayed directly in their comments on art. We also show here that similar differences characterize the comments of high- vs. low-scorers at a younger age, to whom we were not able to give the personality questionnaire.

Emotional reasons given by high-scorers coincide with this pattern also. Liking art with more feeling, with more faithfulness to the spirit of reality and yet with emotionally meaningful stylization rather than photographic literalness, is clearly a better fit with the high-scorer's liking disorder and lack of detail than it is with the older child's liking unity and balance.

One set of reasons correlated with scoring groups forms a quite different pattern. Liking art that seems old or valuable, or art that is recognized, are reasons requiring knowledge of art. High-scorers certainly do seem to know much more about art. These reasons are therefore more probable for the high-scorers than for the low. We would caution, however, against the easy inference that all the reasons given by the high-scorers represent in an equally direct way what they have absorbed by contact with

the cultural stream of words about art. Symmetry, unity, and composition are terms used in discussions of art addressed to children, yet the high-scorers show no special tendency to have picked them up; it is our older children as a whole who have done so, not high-scorers in particular.

Despite considerable agreement between the correlates of age and the correlates of scoring group, then, we conclude that there are important distinctions between the two and that we may characterize them in this way:

(1) Young children like simple and ample stimulation which is kept attractive and unthreatening. As they grow older they can tolerate emotional discomfort and can accept complexity; but they like the complexity to be unified, balanced, and easily understood.

(2) Children (regardless of age) who lack an esthetic approach like clarity and detail, whereas children who have an esthetic approach like disorder, lack of detail, and various complexities of form and emotion which pose a challenge to understanding.

Addendum

We have considered in the text 36 specific categories which showed a significant relation to age, to scoring group, or to sex; one category (suitability to use) which showed a significant relation to scoring group within a single age group; and one category (pleasantness) which showed no significant results. At the bottom of each table is a list of all the variables contributing to the entries. These lists include an additional 45 variables not mentioned in the text, which also show individually no significant results. There remain to list here 11 miscellaneous reasons not represented in any table and --for lack of significant results--not mentioned in the text. Together with percentages of total reasons, they are:

Quality of workmanship, 1.7%. Newness (modernity of work or subject matter), 1.4%. Difficulty of making, 0.8%. Wealth or high status (of subject), 0.5%. Youth (of subject), 0.2%. Old age (of subject), 0.1%. Humility (of subject), 0.1%. Maturity (of work, i.e., a child couldn't have made it), 0.1%. Lack of evocativeness, 0.1%. Childishness, <0.1%. Ease of making, <0.1%.

References

- Barron, F. The psychology of imagination. Sci. Amer., 1958, 199, 150-166.
- Child, I. L. Development of sensitivity to esthetic values. Cooperative Research Project No. 1748, U. S. Office of Education. Mimeographed report.
- Child, I. L. Personality correlates of esthetic judgment in college students. J. Pers., 1965, 33, 476-511.
- Child, I. L., & Iwao, Sumiko. Personality and esthetic sensitivity: Extension of findings to younger age and to different culture. Unpublished manuscript, 1968.
- Francès, R. Variations génétiques et différentielles des critères du jugement pictural. Sciences de l'Art, 1966, 3, 119-135.
- Francès, R., & Voillaume, Huguette. Une composante du jugement pictural: La fidélité de la représentation. Psychol. franç., 1964, 9, 241-256.
- Lark-Horovitz, Betty. On art appreciation of children: I. Preference of picture subjects in general. J. educ. Res., 1937, 31, 118-137.
- Machotka, P. Le développement des critères esthétiques chez l'enfant. Enfance, 1963, 16, 357-379.
- Machotka, P. Aesthetic criteria in childhood: Justifications of preference. Child Devel., 1966, 37, 877-885.

AN ATTEMPT TO ASSESS SPECIFIC ASPECTS OF ESTHETIC VALUE IN VISUAL ART¹

Irvin L. Child

and

Rosaline S. Schwartz

Yale University

We report here an attempt at separate measurement, for research purposes, of specific aspects of esthetic value in visual art. A satisfactory theory of esthetic value, we believe, is not likely to base the worth of a piece of art entirely on the value of its specific aspects. But as a heuristic device, such measurements may be very useful.

Materials

The art we tried to assess consisted of paired slides we have used in other research (e. g., Child, 1964; Child & Schwartz, 1966, 1967). Each pair comprised two works of like kind or subject matter, often similar in style, but differing in esthetic worth according to expert opinion. These pairs, about 900 in number, had previously been shown to school children to obtain expressions of preference.

Procedures

We began by asking two artists (both also art teachers) to try differential assessment of aspects of esthetic worth, using the variables that seemed salient to them, and also by asking a student little experienced in art to attempt ratings on some of those variables. Our final variables and instructions were based on this experience. Ratings made in the preliminary stage are not included in our analysis, however, because the variables used later were differently enough defined that direct

comparison is not appropriate.

Fourteen variables in all were finally used. One applies only to pairs in full color; the other 13 are applicable to all pairs. For each variable, the judge was to compare the two works in a pair and decide which was higher on that variable and to what degree (on a 3-point scale for magnitude of difference in each direction, forming altogether a 6-point scale. The variables, briefly described here, are given in the addendum as they appeared in the judges' manual.

Five of the variables named what may be considered elements of visual art for which the only pertinent question to pose was, Which in the pair is superior in its use of that element? These five are: (1) Form or shape. (2) Line. (3) Movement (including not only directional lines and their interaction but also movement developed from relationships of planes, colors, forms, etc., and even total absence of movement). (4) Lightness and darkness. (5) Color (i. e., hue and saturation, and thus not applicable to the black-and-white pairs).

In another six variables, an aspect of art was again named (though "element" is not in every case an appropriate label). The question now asked was somewhat more objective, less explicitly evaluative, and varied slightly for each aspect: (6) Use of material (In which work is the material used more fully exploited?). (7) Compositional unity (Which has greater unity of composition?). (8) Richness (Which is quantitatively more varied or complex; that is, has more, and more varied, detail?). (9) Sense of space (In which is a sense of space more fully developed?). (10) Tactile quality (Which has more tactile qualities?). (11) Exaggeration and distortion (In which work do distortion and exaggeration play a more effective role?).

Finally, the last three variables, again explicitly evaluative, attempt to get at modes of evaluating the work as a whole rather than at isolable aspects of it. Briefly summarized, these three are: (12) Significance (In which work is an essential and fundamental significance of the subject matter, or of the object itself, so presented that it is clearer to or affects more a skilled and perceptive viewer?). (13) Imaginativeness (In which is the idea or significance expressed with more imaginativeness, so that the viewer's understanding of the significance or idea, or point of view on the idea, is more facilitated or enlarged?). (14) Appropriateness of parts to significance (In which work is there more appropriateness of, a more suitable relationship between, the significance shown and the separate parts--e. g., forms, colors, lines, tactile qualities, etc.--and relations among them, chosen to present this significance?).

We decided to have all 14 variables rated for one pair at a time. The rating was first done by Judge A, who had also been a principal author of the scoring manual. She has no specialized or professional training in visual art (having had training and work experience in literature instead), but years of experience as a viewer and appreciator of art. The rating was subsequently done by Judge B, working with the same manual of instructions; he is a graduate student in history of art interested in problems of esthetic evaluation. Each judge required at least 60 hours to go through the 900 pairs. The first completed the ratings in two periods of intensive work separated by several months. The second spread the judging out into a larger number of shorter periods, finishing about a year after he had started.

Results

We will first consider the 13 variables applicable to all the pairs

before referring to the one applicable only to the colored pairs.

How far did a judge manage to make genuine distinctions among the 13 bases for evaluation? In seeking to answer this question, we began by calculating over the sample of almost 900 pairs the correlation of one judge's ratings on each variable with his ratings on each of the other 12. A factor analysis was then performed on the matrix of intercorrelations to indicate the patterning among variables for that judge. The same process was then followed for the ratings of the other judge. Results are summarized in the first three columns of Table 1.

For Judge B, only one factor had an eigenvalue as large as 1.0--i. e., accounted for at least as much of the variance as a single one of the original variables. This one factor accounted for 59% of the total variance, the next factor accounting for only 6.7%. All 13 variables have substantial loadings on this one large factor, which may appropriately be considered to represent general esthetic evaluation. The size of the loading varies decidedly from one variable to another; this variation is approximately what might be expected for a factor of general evaluation. That is, variables 12-14, by definition closest to general esthetic evaluation, tend to have high loadings, and variables 6-11, by definition the most independent of general evaluation, encompass the lowest loadings. We may conclude that for Judge B all the specific ratings are related to his impression of which work of art is generally superior and the degree of its general superiority, and that the variables do not otherwise fall into appreciable patterns of similarity.

For Judge A the first factor is not quite so dominating. It accounts for 53% of the total variance. For her, there is a second factor with an eigenvalue greater than 1.0, and it accounts for 10.2% of the

Table 1

Specific Aspects of Esthetic Evaluation Rated Independently by Two Judges:
Factor Analysis of each Judge's Ratings, and Interjudge Correlations

Aspect of Esthetic Evaluation	First Factor Loading, Judge B	First Factor Loading, Judge A	Second Factor Loading, Judge A	Correla- tion be- tween 2 Judges' Raw Ratings	Correla- tion be- tween 2 Judges' Residuals
1. Form or Shape	.81	.82	.13	.12	.03
2. Line	.73	.88	.03	.15	.07
3. Movement	.86	.80	.11	.10	.05
4. Lightness-darkness	.88	.81	.10	.12	.06
5. Color				.20	-.02
6. Use of material	.85	.76	.19	.21	.11
7. Compositional unity	.85	.86	.02	.13	.09
8. Richness	.65	-.03	.80	.27	.34
9. Sense of space	.60	.14	.80	.10	.13
10. Tactile quality	.65	.70	.19	.13	.05
11. Exaggeration-distortion	.51	.44	-.07	.33	.36
12. Significance	.88	.86	.10	.17	.00
13. Imaginativeness	.89	.76	.00	.16	.00
14. Appropriateness	.73	.84	.03	.08	.01

total variance. The two variables with high loadings on this second factor, richness and sense of space, have approximately the same intercorrelation for each judge (.34 for Judge B, .33 for Judge A); but whereas for Judge B each has higher correlations with other variables, for Judge A this is the highest correlation either of these variables shows. For Judge A as for Judge B, then, the first factor may be described as one of general esthetic evaluation; the difference between the judges lies in whether richness and sense of space are related to general esthetic evaluation.

The second step was to determine whether the general esthetic evaluation by the two raters agrees. In direction it certainly does; overwhelmingly each rater agrees with the earlier judges about which is the better work in a pair. (Judge A had already taken part in research with these stimuli and knew, for most pairs, which had been judged better. Judge B did not have this knowledge.) But we are concerned here with their quantitative assessment of the degree of superiority or inferiority distinguishing the work experts consider better from the one they consider poorer. This measure was obtained by calculating the factor score for each pair on the first factor emerging from the factor analysis of a judge's ratings on the 13 variables. The correlation between the two judges on this factor score was only .18, not an impressive agreement. It is, however, a degree of agreement which, if maintained throughout a small group of judges would lead to high stability in their consensus.

Could it be that the correlation is this low because one of our judges is incompetent at esthetic evaluation? We have the possibility of answering this question because we have evaluations of the pairs by our original 14 experts. One measure of their evaluation consists of the number who agreed upon which was the better work; this number varies only

from 12 to 14, since we have used only pairs on which at least 12 agreed. With this measure the factor scores of Judge A show a correlation of .21 and those of Judge B a correlation of .26. A second measure is provided by the fact that the original experts also rated on a 3-point scale the degree of superiority of one work over the other; since they could differ on direction, there was thus a 6-point scale altogether. This finer measure, summed over the 14 original raters, is correlated .29 with the factor scores of Judge A and .40 with the factor scores of Judge B. In both instances Judge B shows the higher correlation, and this is the direction of difference we might expect from his being, like the original experts, professionally involved with visual art. But the magnitude of the difference clearly does not suggest that one judge is competent and the other not.

Let us now consider the similarity between the judges in their ratings of specific variables. Here we may include the 14th variable, applicable only to colored pairs. Correlations between the two judges are presented in the fourth column of Table 1; all are positive, and they vary from .08 to .33. There is agreement, then, but is it perhaps solely a result of agreement in general esthetic evaluation? To answer this question we have computed for each judge residuals on every variable--residuals from the regression line for predicting the variable from that judge's general evaluation. This involves some degree of over-correction, because the general evaluation is partly determined by the rating on each variable. We chose this procedure, however, rather than the more laborious one of correcting each variable by use of a measure not influenced by it.

Residuals thus obtained for the two judges were then correlated to measure agreement between judges. The outcome is given in the fifth column

of Table 1. Most of the agreement between our two judges has now vanished. On only two of the variables--richness, and exaggeration and distortion--is there still some substantial agreement. Yet agreement between the judges on the other residuals, though usually very small, is generally highly significant statistically (the standard error of the correlation coefficient, with N approximately 900, is .033). Positive correlation vanishes altogether only for the three variables of most general import (12-14) and for a single one of the more specific variables, that pertaining to color.

The result does not give much hope that the ratings we obtained from just two judges can be usefully related to other facts about the pairs, for their reliability is very low. But it strongly suggests that where comparable ratings can be obtained from many skilled judges, they can be very useful in research. Even our over-corrected residuals show for 9 variables an interjudge correlation of .05 or more. If correlation between single judges were sustained at only this low level of .05, the Spearman-Brown formula would predict for the average of 20 judges a reliability of 0.51, and for the average of 100 judges a reliability of 0.84. Whether correlation among various expert judges would average as high as for our sample of two can be known only by trial. If it did, their average rating of specific aspects of esthetic value would clearly provide measures stable enough to be greatly valuable in investigating the psychology of esthetic response.

Despite the very low reliability of the specific evaluations we obtained, we are exploring their possible correlations with responses of children to works of art. Recurrent computational problems have prevented completion of this exploration in time for inclusion in this report.

References

- Child, I. L. Development of sensitivity to esthetic values. Cooperative Research Project No. 1748 under Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare, 1964.
- Child, I. L., & Schwartz, R. S. Exploring the teaching of art values. Journal of Aesthetic Education, 1966, vol. 1, no. 2, 41-54.
- Child, I. L., & Schwartz, R. S. Exposure to better and poorer art. Journal of Aesthetic Education, 1967 (in press).

Footnote

1. We are extremely grateful to Frederick W. Hornbeck, Vivian Kline, and Joseph Slate for their invaluable contributions to developing the scheme of analysis reported here, and to David Summers for his patient participation in the judging and illuminating comments on the definitions of variables.

Addendum

Here are the definitions of 14 aspects of esthetic evaluation as they were given to the two judges.

1. Form or shape

Which work of art is superior with respect to forms and shapes? So far as possible, try to judge the artistic agreeableness or pleasingness (in a general sense, not restricted to pleasantness or happiness) of the forms and shapes used or represented, and their relation among forms or shapes, apart from their relation to the meaning of the work.

2. Line

Which work of art is superior in its use of line? So far as possible, try to judge the artistic agreeableness or pleasingness (in a general sense, not restricted to pleasantness or happiness) of the lines in themselves, apart from their relation to the meaning of the work. Include not only lines drawn as lines, but also the lines formed or implied by the edges of objects, by the meeting of different colors or textures, etc.

3. Movement

Which work of art is superior with respect to movement? Movement is defined to include directional lines and their interaction, movement developed from relationships of planes, colors, forms, etc., and the total absence of movement to realize the essence of tranquillity. So far as possible, try to judge the artistic agreeableness or pleasingness (in a general sense, not restricted to pleasantness or happiness) of the movement, apart from its relation to the meaning of the work.

4. Lightness-darkness

Which work of art is superior in its use of variations in lightness and darkness? Lightness-darkness is used here to include several aspects which are partially independent but are placed together for judging: the lightness-darkness of the separate parts of the work, this as modified by interaction with other parts, and the sense of illumination vs. shadow developed in representational works and in some abstract paintings. Weight these several aspects as seems appropriate for the particular pair. Where the works of art are objects (such as statues or vases) which do not occupy the entire photographed area, concentrate on the light-dark relationships within the objects themselves rather than on relationship to background. So far as possible, try to judge the artistic agreeableness or pleasingness (in a general sense, not restricted to pleasantness or happiness) of the variations in lightness-darkness and the way they are employed in relation to each other, apart from their relation to the meaning of the work.

5. Color

Which work of art is superior in its use of color? Color here refers to the hues used and their saturation (lightness-darkness will be considered as a separate variable). So far as possible try to judge the artistic agreeableness or pleasingness (in a general sense, not restricted to pleasantness or happiness) of the colors used and the way they are combined, apart from their relation to the meaning of the work.

6. Use of material

In which of the two works of art is the material used to form the object or represent the subject matter--charcoal, water colors, clay, paint, marble, etc.--more fully exploited to develop those qualities by which we recognize and understand all those objects and forms that are shown and their significance?

7. Compositional unity

Which work of art has greater compositional unity? In which work of art are the separate components (parts of the object or of the subject matter represented, forms, colors, lines, movements, light and dark, etc.) more suitably placed in relation to all other parts to develop a visual unity; *i. e.*, which work of art more tightly integrates all parts of the figures and the ground, causing each element to contribute to one whole?

8. Richness

Which work of art is quantitatively more varied or complex, that is to say has more, and more varied, detail to be absorbed visually? The variation, complexity, or richness can reside in form, color, line, movement, light and dark, tactile qualities, depth, and all or any combinations of these.

9. Sense of space

In which work of art is a sense of space more fully developed? Sense of space is not limited to depth, nor to linear perspective, but may include any means (size, color, relationships of forms, etc.) that may have been used to produce the feeling of space.

10. Tactile quality

Which work of art has more tactile qualities, *i. e.*, arouses more actively and fully in a sensitive and skilled viewer such things as feelings of roundness, smoothness, heaviness, wetness, heat, etc., and their opposites?

11. Exaggeration and distortion

In which work of art do distortion and exaggeration of color, light, shadow, line, form, etc. play a more effective role in revealing to the viewer the significance of what is shown?

12. Significance

In which work of art is an essential and fundamental significance of the subject matter (or the object, in the case of "useful" or non-representational art) so presented that it is clearer to or affects more a skilled and perceptive viewer? By significance we mean an idea of or about, or a useful or important or valuable way of seeing, the object or the subject matter. Examples of significance are an interpretation of the basic character of a person, a view of the world or an emotional state expressed in an abstract painting, the sunniness of a sunflower, the containing power of a pitcher, the total exultation in the moment of the Virgin's coronation, the "seatiness" of a chair, the lifelessness of death, the fecundity of springtime.

In an extreme instance, there may be a contrast between a highly

significant work and one which seems to express nothing significant about the subject it represents or the object it is. In less extreme instances, the two works need to be compared with respect to the importance of the idea conveyed, its appropriateness, and its clarity for a skilled and perceptive viewer.

13. Imaginativeness

In which of the two works of art is the idea or significance expressed with more imaginativeness, so that the viewer's understanding of the significance or idea, or point of view on the idea, is more facilitated or enlarged?

Here we are concerned with the visual artist's "imagery," his "metaphor" or "simile," developed by his selection of details to represent, demonstrate, or illustrate the significance of his subjects or objects, and by his use of the formal elements of the visual arts.

Here are some possible examples (so far as they can be stated in words) of imagery that would be considered highly imaginative; the examples have been chosen, too, to illustrate that very diverse imagery may be equally appropriate for the same basic idea.

The spiritual quality of a religious building may be expressed by stressing long, vertical lines and by delicate tracery, thus suggesting the direction and directness of the relation between man and his god as well as the richness and delicacy of that relationship; or it can be expressed by horizontal lines, small in scale, and by the use of blazing white, suggesting that from the beginning, and before all else (before color), the spiritual is present.

A quality of heaven may be shown by a lamb and a lion joyfully dancing together, or by innumerable angels with innocent faces walking weightlessly on meadows of fresh colored flowers.

The "seatiness" of a chair can be developed by a bare horizontal solidly resting on straight legs, or it may be developed by a curving mass that resembles the seated human figure.

The loneliness of man may be suggested by blank, undetailed shapes contoured to resemble the human body, or by a highly developed, detailed, quiet and sorrowful human face surrounded by brilliantly colored details that represent a highly mobile, excited scene.

The degree to which the imagery is imaginative can be judged by the extent to which it is concrete and specific in illustrating the significance, by the way it avoids or supplements ordinary literal description. Aptness of metaphor must enter into the judgment. Unusualness of metaphor is not in itself what is being judged; what is being judged is how much the metaphor sharpens understanding or remembrance of the significance.

This variable should be judged even for works of a completely abstract and non-functional character.

14. Appropriateness of parts to significance

In which work of art is there more appropriateness of, a more suitable relationship between, the significance shown and the separate parts chosen to present the significance? By separate parts are meant the parts of the object or of the subject matter represented, the forms, colors, lines, tactile qualities, depth, and the relationships among these.

Determinants of Children's Art Preferences and of their Agreement
with Expert Esthetic Judgment¹

Irvin L. Child, Yale University

(A paper prepared for presentation at the Second International Colloquium
on Empirical Aesthetics, San Marino, September 17, 1966)

We begin with a fact, and seek to understand it better.

The fact we begin with is this: Children in elementary schools tend to prefer art which experts consider esthetically poorer, rather than art experts consider better. As children's preferences change during the years of secondary school, this disagreement with experts decreases.

This fact was long ago well established for British children by Margaret Bulley, who also provided confirmatory evidence for groups of children in France, Cyprus, and Japan.² In our research it is again confirmed for children in the United States. Since the explanations I will suggest arise from the data I have obtained, I must begin by reporting more specifically the way in which those data confirm the general finding.

We have shown to children over 900 pairs of projected slides. Each pair consists of two works of art, similar in type or subject matter, and usually similar in style, differing in esthetic merit according to the person--usually a present or former university art student--who made up the pair. The direction of the difference was confirmed by 12 or more out of 14 experts--again, mostly present or former university art students--who were asked to judge independently which work in each pair was better esthetically.

Elementary schools in the United States include grades 1 through 6.

In the lowest two grades, many children respond carelessly to a task as complicated as ours; so I will simplify by reporting elementary-school data only from grades 3 through 6, with children aged mostly between 8 and 11 years. We studied several elementary schools, but I will confine myself to the one whose children come from families most like those from which our older children come. In the elementary school all the pairs were shown, in a series of sessions, to about 50 children in each grade. In secondary schools, which include grades 7 through 12, with children aged mostly between 12 and 17 years, we had several thousand children available; in general, each took part in only one session at which 130 pairs were shown, with different sets of pairs being shown to various classrooms of each grade. The children were not told anything about how or why the pairs were made up, except the fact that we were interested in learning about children's art preferences. The examiner asked that each child look carefully at the pictures in each pair and then indicate which work he liked better.

To what extent do the children prefer that work in each pair which the experts consider to be the better art? In the elementary schools, the agreement with experts averages about 40%, indicating a definite tendency to prefer the work the experts consider poorer. Through the years of elementary school there is no consistent change in this percentage. In the six grades of secondary school, on the other hand, we do find regular change from year to year. The percentage of agreement rises year by year until it finally approaches but does not quite reach 50%.

There is the main fact with which we start. Now how can we hope to find its meaning? One hope lies in studying variations in the agreement with experts.

At any age, for example, there are variations among individual persons in the extent of agreement with experts. At the first International Colloquium on Empirical Aesthetics, I reported research on college students focussed on the characteristics of individuals who tend to agree with experts and those who do not.³ If preferences are generally related to personality, an account of the personalities of people whose preferences agree with experts might make clear the personal meaning of art for them and thus indirectly get at the distinctive psychological function of the esthetic approach to art. We have been able to use this approach with secondary-school students, too, and have confirmed in them that certain personality characteristics--independence of judgment and tolerance of complexity--are associated with preferences which agree with those of experts.⁴

But there is another important variation in agreement with experts. Agreement not only varies from person to person; it also varies from pair to pair. Here is another opportunity to learn something about the bases for agreement, and this is the approach I want to discuss.

I will begin with some examples, drawn from the black-and-white pairs which formed about half of our material. First, two drawings of a woman's head: the one experts consider better is by the 19th-century artist John Singer Sargent, the other is an illustration from a manual on how to teach yourself to draw. The Sargent is preferred by only 10% of our elementary-school children, and by 38% of our high-school children; the results show the typical change, though with lower-than-average agreement with experts at each age. Next, consider two wash drawings of a man's head--a self-portrait by Rembrandt and a recent portrait of Leonardo prepared for an advertisement. The results in elementary school are typical, with 42% preferring the Rembrandt; in high school we find practically no change, with 40% preferring the Rembrandt. Finally, consider two still-lives where the

one by Vlaminck has by our judges been considered better than the one by Pignon. Here the agreement with experts is very frequent in the elementary school, 72%, but it declines to 52% in the high school.

What do the works of art children prefer have in common? What do the works of art they reject have in common? Previous research on the bases for art preferences in children, and on changes in those bases with increasing age, can be studied for possible answers. Following a long tradition of concern with art education, systematic research that is highly relevant has been done in recent years by Subes, by Zavalloni and Giordani, by Machotka, by Francès and Voillaume, and by Francès as reported at the first International Colloquium last year.⁵ And research on rated characteristics of art, done by Peel, by Pickford, by Cardinet, and by Gordon, even though not concerned with children's response to art, is also highly useful.⁶ Especially valuable is Peel's idea of measuring for the same stimuli their qualities as objects and their tendency to be liked, and then determining the relation between these two sets of variables. Instead, however, of dealing with the qualities of each single work and response to it, we deal always with the qualities of contrast presented by a pair, and response to the pair. To draw up a list of 24 characteristics that could be reasonably well measured and seemed relevant to the contrasts presented by our pairs, we drew upon all past research, accepted suggestions made in interviews by a number of the children whose preferences we had studied, and were guided by our intuition and a bit of theory. Naturally, then, this research at several points is confirming the findings of these earlier studies; in this brief account, however, I will not try to discuss in detail all the agreements and the occasional disagreements.

As I have stated, our unit stimulus is the pair. We have for each pair a measure of the proportion of children who prefer the work of art which experts consider better esthetically. We have this measure separately for boys and for girls at each school grade, but in general the two sexes yield very similar results. In this respect, our findings agree with those reported last year by Francès on children's stated reasons for preference. For today's presentation I have averaged results obtained from boys and girls. I have combined elementary grades 3 through 6. I have also put together the highest three secondary-school grades, 10 through 12, or what we often refer to as the senior high school, to contrast this extreme of development with what is found in the elementary school. For one purpose where larger numbers are needed, I have joined with these the other three secondary grades, 7 through 9, often referred to in the United States as junior high school.

The characteristics of the pairs, to be related to the preferences of school children, were assessed by having them rated. The characteristics I am reporting on here are objective rather than evaluative. That is, they do not pertain directly to the relative esthetic merit of the two works and can be judged without any reference to esthetic merit. To help ensure that they were so rated, indeed, we asked our judges to compare the two pictures as they saw them projected and to estimate whether the one on the left or the one on the right was higher on the characteristic being rated, and by how many points on a 3-point scale. The rating was thus of the extent of differentiation within the pair on a particular characteristic. In working with the data after the ratings had been completed, we began by adjusting the ratings so their direction would have a constant meaning--i. e., so

that a positive rating, +1, +2, or +3, would mean that the esthetically better work had more of this characteristic, and a negative rating, -1, -2, or -3, would mean it had less. If the two works did not differ noticeably, the rating was zero.

Each characteristic was rated by two judges, adults, varying widely in acquaintance with art. We asked them to try to take a naive attitude and to judge the characteristics either as objectively present (where that was pertinent) or as they would probably appear to a viewer with no specialized knowledge about art. The independent ratings of a variable by two different judges were correlated with each other. Since we were going to use the average of the two judges' ratings, it was appropriate to apply the Spearman-Brown correction to these coefficients; with the exception of one variable with a corrected value of .25, the result varies from .41 up to .82. The agreement is generally sufficient to be useful, though it is clearly not uniformly so high as would be desirable; only the cost and tedium of making the ratings leads us to be for the moment satisfied with this degree of reliability.

Of the 24 variables rated, 17 yield results which seem clearly relevant to our problem. These 17 fall into four groups; I will discuss each group in turn.

SENSORY QUALITIES

The characteristics I consider as sensory qualities are lightness, light-dark contrast, warmth of hue, hue variation, saturation, and saturation variation. Three of these are defined by the three basic dimensions of color, and the other three refer to variation within a single dimension; all six are pertinent to change in preference during the school years.

Four of these characteristics, of course, are not applicable to the pairs made up of black-and-white pictures; such pairs were omitted from all calculations with those variables.

First, we have the relative lightness of the two pictures, the extent to which the total or average amount of light perceived in the better work is greater or less than that perceived in the poorer work. This characteristic has almost no relevance for the preferences of elementary-school children, as indicated by the correlation of $-.03$ entered in the first cell of Table 1 (page 22). But it does have a negative significance for the preferences of high-school students as indicated by the correlation of $-.11$ in column (2). The difference between these two correlations means that this characteristic is relevant to changes in preference from the earlier to the later ages. This fact is separately represented by the correlation $-.08$ in column (3), which indicates the extent to which greater lightness of the better work is correlated with its being more preferred by high-school than by elementary-school children. Since the coefficient is negative, the tendency is in the opposite direction.

Let us consider this variable fully, as an example of how the table may be read in detail for the subsequent variables as well. Columns (4) and (5) provide us with information about whether the relevance of this variable for secondary-school preference is fairly homogeneous through the secondary-school population or varies markedly according to whether individuals have come to share appreciably the esthetic preferences of experts. To have large enough samples, we now include the entire six grades of secondary school. For each set of pairs shown together we have selected, from the hundreds of students who responded to it, 50 of the highest scorers--that is, 50 who showed the greatest tendency to agree with experts.

Generally, they agreed with experts in at least 55% of their preferences. We selected them so as to be well spread out over the six grades, and then selected low scorers with the same grade distribution. For each pair we then determined the number of low scorers who preferred the work experts consider better, and the number of high scorers who did so. These two quantities could then each be correlated with the rating of relative lightness; the results appear in the next two columns of the table. We see a difference. The low scorers are influenced by relative lightness in about the same way as the elementary-school children ($-.06$ as compared with $-.03$), whereas the high scorers carry further than high-school students generally the tendency of the latter to be negatively influenced by lightness ($-.16$ as compared with $-.11$).

Finally, in the last column of Table 1 we find an entry which shows to what extent the differentiation found by our raters on this characteristic tends to be in a consistent direction. An entry of $+1.0$ here would mean that on the average each rater reported the better work to be lighter by half a point on the 3-point scale they used to judge the amount of difference. The entry here is $-.6$, and this means that the opposite tendency was found consistently; for a single rater, the poorer work was found to average three-tenths of a point higher in lightness. This indicates that changes in relation of preferences to relative lightness will be relevant to changes in their relation to expert judgment.

Now let us consider as a group the other five sensory qualities. We find here decided uniformity. The correlations with elementary-school preference are all positive; at that age, children tend to prefer the picture with more light-dark contrast, warmer colors, more variation of hue, more saturated colors, and greater variation in saturation. In a rough

way, and with some inaccuracy, we could generalize that the younger children prefer more colorful art. In the high-school students these correlations have all disappeared: these sensory variables have very little relation to their preferences; some coefficients are positive and some negative, but all are very small. The shift is expressed also by the uniformly negative correlation between these variables and the change in preference from elementary school to high school. When we consider separately the low scorers and high scorers in the secondary schools, we find that the low scorers still share some of the younger children's preference for colorfulness, though in reduced degree. The high scorers, on the other hand, show definite correlations in the other direction. Except for light-dark contrast, finally, we see in the last column that these changes are relevant to agreement with expert judgment, for the works experts consider better do seem to be consistently the less colorful on the average.

These findings do not clearly establish an interpretation. Discussion of them might well be lengthy, especially in relation to research on preference in response to single patches of color, such as has been conducted by Eysenck, Granger, and Guilford.⁷ I will sketch an interpretation which for the sake of brevity will avoid that issue.

In viewing works of art, it appears that young children enjoy being assaulted or stimulated by warm colors, by saturated colors, by wide variation in hue, saturation, and lightness. Perhaps this is because of a biologically determined pleasure-value in such stimulation. Perhaps it is because the resulting arousal is pleasant. Perhaps it is specifically because in the confining and restraining situation of the school, arousal rather than relaxation is needed to move them toward optimal arousal level. Whatever the source of the connection in childhood, it has disappeared in

the average high-school student. For the present I am inclined to regard this as a consequence of the increasing complexity of the growing person and hence of his reaction to pictures. Able to respond more fully to a variety of other aspects of art, his choice is little influenced by this purely sensory aspect. (This interpretation is consistent with Subes' report⁸ that his subjects, who likewise showed with greater age lessened preference for vivid and varied colors in paintings, showed no change in preference for colorfulness in geometric assemblies which presumably lacked esthetic appeal.) That the experts' choices tend to be lower on most of these variables than are the works they consider poorer, I would suggest, may be a product of the fact that art may please in various ways. Art that does not please by the esthetic criteria used by experts may please by offering the sensory qualities we are considering here. A fine work of art can have all these qualities, but need not; a poorer work of art may need these qualities in order to attract any attention at all. (And in general, the works in our pairs need to have attracted some attention to have been published or in some way reproduced in order to be available for our use.) The negative correlation of our high scorers' preferences with these qualities would, on this interpretation, be an incidental by-product rather than an indication that they are avoiding these qualities.

EMOTIONAL QUALITIES

Under emotional qualities I will group happiness, sentimentality, and emotionality. The ratings of happiness were largely influenced by the subject matter portrayed; a portrait of a smiling child would be rated somewhat happier than one that lacked expression, and much happier than one expressing gloom. But the ratings were also influenced by the emotional

meaning of the colors, lines, etc., and even a pair of non-representational paintings could be differentiated on this variable.

Happiness has an important influence on the preferences of elementary-school children, with a correlation of .32. In high school there is still a substantial tendency to prefer the happier work, although the fact of change is indicated by the correlation of $-.17$ between the happiness ratings and change in preference. The low scorers in secondary school are still about as much influenced by happiness as are the younger children; but in the high scorers any correlation of preference with happiness has completely disappeared. The differentiation between better and poorer work is more consistent on this characteristic than on any of the other 23; the mean deviation from zero, appearing in the last column of the table, is $-.9$.

Results for sentimentality are similar though not so extreme. Sentimentality was defined as "the display of superficial emotion of a sort, perhaps, that could be easily aroused and easily suppressed--emotion that might be felt to be excessive and, at any rate, is expressed in a manner not inviting the control or discipline that might be aroused by a more rational or intellectual presentation. Sentimental expression of emotion may convey a suggestion that what is to be enjoyed is the fact that an emotion is occurring; sentimental expression may give little emphasis to which particular emotion it is, or to laying the groundwork for experiencing that emotion with some depth of context and understanding." It is surprising that so complicated a concept could be applied with good inter-judge agreement, but we found that it was. Like happiness, sentimentality declines in importance as a positive determiner of preference, from elementary school to high school, and is least important in the high scorers. Since the raters

find sentimentality to predominate more often in the poorer art, this variable too helps explain the change in agreement with experts.

The ratings of emotionality were made in response to the following instructions: "Which picture has more intense emotional or dramatic quality? Here we are not concerned with the kind of emotion--happiness, sadness, etc.--nor whether the emotion is obvious or subtle. Effort should be directed toward judging the difference in intensity of emotional or dramatic quality a viewer would be likely to see in the pictures. (You should have in mind here a viewer who has no background of knowledge about the scenes or the painters, but who does understand what he sees in front of him as fully as its own characteristics make possible.)" Here we felt the concept required an exception to the notion that the rater should think of a viewer naive in every way; an understanding of the picture was thought necessary. We had expected that emotionality would yield results opposite to those obtained with the other variables in this group. Our expectation that emotionality would be rated higher in the esthetically superior works was to some extent confirmed, with a mean rating of .4. We had also expected that emotionality would be somewhat negatively correlated with the younger children's preference and perhaps positively with the older. There was a movement in this direction, though it began with almost no correlation at all in the younger children ($-.01$) and the correlation for the high school reached only .10. But little can be said with certainty about the meaning of these results because this is the one variable for which the inter-judge agreement was so low that we can have no confidence in what the ratings mean.

For the first two emotional qualities, however, we can make a more definite inference. Portrayal of happiness and of superficial solution of

potential emotional problems appeals to younger children. It seems to be a simple, direct form of gratification. The liking for happiness might be thought of as akin to the liking for warm and saturated colors. Liking for sentimentality might be thought of rather as an avoidance of the high degree of arousal that might result from contemplation of pictures presenting a more genuine view of sorrow or tragedy. In contrast with sheer color stimulation, emotional stimulation might be thought of as easily being in excess, with sentimentality serving to keep arousal down at a desirable level. On this hypothesis, the high-school student has learned to cope with emotional arousal. Even in the high-scoring students, though, there seems to be some trace of the liking for simple happiness and sentimentality, mingled with a toleration of unpleasant or unresolved emotion, to produce the near-zero correlation between their preferences and these variables despite the decided tendency for the better works to be rated less happy and less sentimental.

ABUNDANCE

The set of characteristics I group as abundance consists of completeness, abundance of skill (difficulty of making), and abundance of detail (quantitative fullness). All three give closely similar results.

The variable found to be most closely related to preference was the one we called "completeness." Instructions for raters defined it as follows: "Apart from considerations judged under 'realism of representation,' to what extent does one picture or work seem likely to give a naive viewer the impression of being more complete or finished? Include here such features as whether a person is cut off by the edge of the canvas, whether a picture looks as though it had been left half done, whether the wood surface

potential emotional problems appeals to younger children. It seems to be a simple, direct form of gratification. The liking for happiness might be thought of as akin to the liking for warm and saturated colors. Liking for sentimentality might be thought of rather as an avoidance of the high degree of arousal that might result from contemplation of pictures presenting a more genuine view of sorrow or tragedy. In contrast with sheer color stimulation, emotional stimulation might be thought of as easily being in excess, with sentimentality serving to keep arousal down at a desirable level. On this hypothesis, the high-school student has learned to cope with emotional arousal. Even in the high-scoring students, though, there seems to be some trace of the liking for simple happiness and sentimentality, mingled with a toleration of unpleasant or unresolved emotion, to produce the near-zero correlation between their preferences and these variables despite the decided tendency for the better works to be rated less happy and less sentimental.

ABUNDANCE

The set of characteristics I group as abundance consists of completeness, abundance of skill (difficulty of making), and abundance of detail (quantitative fullness). All three give closely similar results.

The variable found to be most closely related to preference was the one we called "completeness." Instructions for raters defined it as follows: "Apart from considerations judged under 'realism of representation,' to what extent does one picture or work seem likely to give a naive viewer the impression of being more complete or finished? Include here such features as whether a person is cut off by the edge of the canvas, whether a picture looks as though it had been left half done, whether the wood surface

of a table looks as though it had been finished or left rough, etc. Incompleteness resulting from wear or breakage should be treated in the same way as incompleteness presumably present from the beginning."

Second in importance was the variable of difficulty in making. Here the rater was asked to judge how hard, in the probable opinion of naive viewers, it would be to produce the work, how much skill and labor would be required.

Third was a rating of quantitative fullness: "Which picture has more in it? No consideration should be given to making an esthetic evaluation of the simplicity or richness or complexity, to estimating the unity or lack of unity of what is shown. All that is wanted is a rating of general impression of how much or how many things or features (colors, shapes, textures, etc.) are presented."

Analysis of the results shows that each of these characteristics strongly influenced the preferences of elementary-school children. In high school, the general influence is only slightly lessened. Among secondary-school students, the low scorers still exhibit preferences which are almost as strongly determined by these variables as are the preferences of the elementary-school population; the high scorers, however, show a greatly diminished influence of these variables or, in the instance of quantitative fullness, none at all. Despite the generally predominant influence of these variables, then, some of the secondary-school students are becoming free of it. The differentiation within the pairs, shown in the last column of the table, is in a uniform direction, the poorer work being higher in abundance--though for difficulty in making, the tendency is small (yet still highly significant statistically).

That young children should apply to art these quantitative standards

is not surprising, for they apply them to other fields. A large ice-cream cone or a large steak is more satisfying than a small one, even if it cannot be devoured. The dazzling goal-kicker is more likely to be a hero than his team-mate whose skill is less spectacular. A bicycle equipped with all appurtenances is more desirable than its stripped-down basic model. A bigger picture, a more complete one, one more filled with details, one that seems to have required more skill or time to make, should be better for this quantitative greatness. That an extension to art of these standards should already begin to disappear in secondary-school students is perhaps more surprising than its presence in younger children.

But why should our raters have found the poorer works to meet these quantitative standards more fully than the better works? Certainly a fine work of art can be quite complete, very full, and quite obviously hard to make. But perhaps it needs to be very fine indeed to assimilate these qualities; completeness, fullness, and obvious difficulty of production may militate against high quality. I suspect, though, that there may be operating here a process parallel to the one I conjectured for the sensory qualities: a work lacking in esthetic appeal may be constructed to appeal through abundance instead.

SIMPLIFICATION

The fourth set of variables--realism, conventionality, clarity, lack of ambiguity, and sharpness of edges--I group together under the label of simplification of experience, for they seem to have in common a tendency to maintain or to extend simplicity.

For realism of representation, results are somewhat distinctive in that its general influence seems to be as great in high school as in

elementary school. In judging realism, raters were asked, "Which work has a more faithful, literal, or realistic representation of the subject matter portrayed?" (Where there was no subject matter, a rating of zero difference was given.) This characteristic of pairs is highly correlated with response at all age levels we have tested; it has a correlation of .40 with elementary-school preference, and a slightly higher correlation of .44 with high-school preference. Consequently, it is very little related to change in preference, despite the fact that it is a variable which tends to differentiate consistently between the better and poorer works in a pair. At all ages the average child strongly prefers the more realistic work, even though the esthetically better work is usually less realistic. In the high scorers among the secondary-school students, however, change may be seen; realism remains a positive influence with them only to the extent indicated by the much lower coefficient of .24.

The remaining four variables in this group show very homogeneous results. For conventionality, the rater judged which work was likely to appear more "familiar, usual, or conventional to the naive Connecticut viewer of recent years, whose acquaintance with art comes primarily from household, church, school, and mass media." Unlike our other variables, this one was thus intended to be judged in relation to the specific experience of the kinds of people whose preferences we were studying. I place it under the heading of "simplification" not because of believing that the more conventional art is itself simple, but because of believing that art which conforms to expectation promotes a simplification of total experience. Unconventional art complicates experience by introducing novelty and challenge.

Clarity and lack of ambiguity are closely related to each other.

The distinction we intend is that clarity refers only to the visual qualities of the work, whereas lack of ambiguity refers to ease of understanding and absence of multiple possibilities of meaning. Some of Dali's pictures show the possibility of combining high clarity with a maximum of ambiguity.

Last among these variables is the sharpness of edges within the work. This characteristic is a specific stylistic element which can be rated separately and may be thought of as contributing to ease of perception and understanding.

These last four variables have essentially the same results. They are highly related to preference in elementary-school children and only a little less so in high-school children. (The change is greater for sharpness of edges than for the other characteristics.) The change in preference from the earlier age to the later is substantially correlated with these kinds of differentiation; the more a work is distinguished from its mate in any of these aspects of simplicity, the more is its choice likely to decrease from elementary-school to high-school years. The low scorers in secondary school show even greater influence of all these characteristics than does the total population in elementary school; in the high scorers, on the other hand, the influence of these characteristics is greatly reduced, and for sharpness of edges it is even reversed. Finally, all of these variables show a considerable tendency to be ^{lower} ~~higher~~ in the better work than in the poorer, so that the changes I have just described are highly relevant to increasing tendency to choose the works considered better.

The five characteristics considered here, then, contribute to understanding the fact with which we started. Elementary-school children like art that reaffirms and exaggerates the simplicity and understandability of the world in which they live. They tend to like art which faithfully

mirrors reality (though in this one respect the peak of development comes later,⁹ in junior high school). They prefer that it be the kind of art they have already encountered in their daily life. They like art which is clearly perceivable and leaves no uncertainties about what it means. A style which uses sharp, crisp edges contributes to their liking. All of these characteristics are compatible with fine art, but a fine art can perhaps not be limited to them alone and remain fine. Certainly in our day the taste of art experts calls for a stirring up, a challenge to understanding, rather than uniform simplification. The American literary theorist, Peckham, has indeed argued in his recent book, Man's Rage for Chaos, that a steady call of this sort is involved in the whole history of the arts and is the force responsible for repeated stylistic change. The function of art for man, he suggests, lies in its offering novelty of experience and thus giving practice in coping with change or discontinuity. Clearly such an orientation is what distinguishes our high scorers from the low scorers; they have become largely immune to the siren-call of simplicity, and are beginning to tolerate departure from realism and from convention, to accept ambiguity, lack of clarity, and fuzziness. Perhaps in time the students who appear in our secondary-school sample as the high scorers may develop a positive liking for these complicating and challenging characteristics.

DISCUSSION

In seeking to understand an initial fact, then, we have demonstrated a number of additional facts that may help us. These new facts, however, do not establish beyond doubt any one pattern of understanding. They only

suggest one. I will present what they suggest to me.

I would give central place to the element of simplification. The esthetic function of art seems to involve the challenge and the mastery of complexity. Expert judges assess art partly by reference to its adequacy for serving this function. Fine art can also serve simpler needs. But to be able to appreciate fine art consistently and with an esthetic point of view, it is necessary to tolerate or even want the challenge of complexity. Younger children usually cannot absorb or accept complexity. They like art in which they find the security of stability, of affirmation that they live in a simple world they can understand. They value art which affirms happiness or gives occasion for other, even unhappy, emotions in a setting which offers bland simplicity, containment, and only superficial disturbance. This blandness can be relieved by the basically pleasant stimulating value of warm, saturated colors and of color variation. They value, too, art that suggests abundance through its wealth of content or through the skill exerted to produce it, and here too is security and the simplicity of completeness. As children grow older, some of them begin to tolerate complexity and even to seek it out. They will look for more challenging emotional experience in art and will tolerate incompleteness and lack of abundance. The sensory stimulation of strong color is no longer necessary and may even distract from what they find interesting in art.

Our discussion follows Machotka in that we view the development of esthetic preferences in children as a part of their general cognitive development. Our findings highlight the differences among various children in the development of esthetic preferences. Many still have in high school the same preferences, and apparently the same reasons for them, that

characterize the general elementary-school population. We may conjecture from what is known about variations in adult taste that the same thing remains true through adulthood. Further elucidation of the development of children's response to art may depend, therefore, on attention to variations as well as to uniformities in general cognitive development.

FOOTNOTES

1. The research reported here was done under Cooperative Research Contracts 1748 and 2840 with the United States Office of Education.
2. Margaret H. Bulley, An enquiry as to aesthetic judgments of children, British Journal of Educational Psychology, 4(1934), pp. 162-182.
Also Margaret H. Bulley, Art and counterfeit, London, 1925, p. 92.
3. I. L. Child, Des études psychologiques qui se rapportent à l'objectivité de la valeur esthétique, Sciences de l'Art, (1966). Some of these results are also reported in I. L. Child, Personality correlates of esthetic judgment in college students, Journal of Personality, 33 (1965), pp. 476-511.
4. I. L. Child and Sumiko Iwao, Personality and esthetic sensitivity: extension of findings to younger age and different culture. (In preparation)
5. J. Subes, La sensibilité de l'enfant à l'art pictural, Enfance, 8(1955), pp. 345-368. R. Zavalloni and N. Giordani, Ricerca sulla sensibilità estetica nell'età evolutiva, Problemi della Pedagogia, 2 (1958), pp. 904-919. P. Machotka, Le développement des critères esthétiques chez l'enfant, Enfance, 16(1963), pp. 357-379.
R. Francès and H. Voillaume, Une composante du jugement pictural:

la fidélité de la représentation, Psychologie française, 9(1964), pp. 241-256. R. Francès, Variations génétiques et différentielles des critères du jugement pictural, Sciences de l'Art, (1966), pp. 31-45.

6. E. A. Peel, On identifying aesthetic types, British Journal of Psychology, 35(1944), pp. 61-69. R. W. Pickford, "Aesthetic" and "technical" factors in artistic appreciation, British Journal of Psychology, 38(1948), pp. 135-141. J. Cardinet, Préférences esthétiques et personnalité, Année psychologique, 58(1958), pp. 45-69. D. A. Gordon, Methodology in the study of art evaluation, Journal of Aesthetics and Art Criticism, 10(1952), pp. 338-352. R. H. Knapp also generously provided us with information about ratings used in a highly pertinent unpublished study by himself and T. C. Ridout.
7. H. J. Eysenck, A critical and experimental study of colour preferences, American Journal of Psychology, 54(1941), pp. 385-394. G. W. Granger, An experimental study of colour preferences, Journal of General Psychology, 52(1955), pp. 3-20. J. P. Guilford, A study in psychodynamics, Psychometrika, 4(1939), pp. 1-23.
8. Op. cit., p. 347.
9. This fact should come as no surprise to readers of Machotka, op. cit., and Francès and Voillaume, op. cit.

TABLE 1 (p. 22)

	(1) Correlation with Elementary- School Preference	(2) Correlation with High- School Preference	(3) Correlation with Difference in Average Preference, High-Elem.	(4) Correlation with Low Scorers' Preference	(5) Correlation with High Scorers' Preference	(6) Mean Rating
SENSORY QUALITIES						
Lightness	-.03	-.11	-.08	-.06	-.16	-.6
Light-dark contrast	.11	-.03	-.11	.14	-.14	.0
Warmth of hue	.08	-.01	-.07	.00	-.06	-.4
Hue variation	.10	-.04	-.13	.02	-.18	-.7
Saturation	.21	.03	-.19	.15	-.17	-.5
Saturation variation	.17	.05	-.14	.14	-.12	-.5
EMOTIONAL QUALITIES						
Happiness	.32	.18	-.17	.30	-.03	-.9
Sentimentality	.21	.15	-.10	.19	.07	-.6
Emotionality	-.01	.10	.10	.06	.08	.4
ABUNDANCE						
Completeness	.49	.40	-.20	.48	.15	-.6
Difficulty of making	.32	.28	-.09	.28	.17	-.1
Quantitative fullness	.25	.13	-.17	.18	-.01	-.4
SIMPLIFICATION						
Realism	.40	.44	-.05	.48	.24	-.5
Conventionality	.39	.35	-.12	.42	.12	-.6
Clarity	.38	.30	-.16	.42	-.01	-.6
Lack of ambiguity	.35	.32	-.10	.40	.11	-.5
Sharpness of edges	.29	.16	-.18	.30	-.11	-.6
OTHER VARIABLES						
Strength	.03	.09	.04	.04	.06	.2
Activity	.13	.12	-.02	.10	.10	-.1
Asymmetry	-.01	.04	.04	.05	.03	.1
Masculinity	-.04	.03	.09	-.03	.05	.1
Curvature	-.01	.06	.07	.02	.09	.2
Shininess	.19	.22	.01	.24	.13	-.2
Close-up view	.01	.05	.06	.04	.00	.0

AGE AND SEX DIFFERENCES IN CHILDREN'S COLOR PREFERENCES

Irvin L. Child, Jens A. Hansen, and Frederick W. Hornbeck

Yale University

Children are often said to have hue preferences markedly different from those of adults. A recent manual on color, for example, says, "Children's color preferences develop and shift with age, showing a tendency to move from warm to cool colors with increasing years" (Burnham, Hanes, & Bartleson, 1963, p. 212). Careful study of such changes has, however, hardly started. There are indeed some quantitative findings on color preferences in relation to age. Beebe-Center (1932, pp. 300-307), summarizing from studies reported before 1932, concluded that "From the third to the fifteenth year, preference for warm colors over cold ones gradually disappears." "Warm" and "cold" are not very precise terms for color, and judgments of warm-cold as part of the connotative meaning of color may not be influenced only by hue (Wright, 1962). But Beebe-Center seems clearly to have intended reference to hue alone. Little new evidence has been added since then, though Bjersted (1960) has recently reported for the Pfister Color-Pyramid Technique that children tend with increasing age to move from warm to cool colors.

A difficulty with most of the developmental studies of color preference thus far is that the stimuli have either been selected so casually or have been so complex (as in Bjersted's study) that we cannot be sure what dimensions of color experience have influenced preference. Where hue has been the intended variable, it has often been confounded with value (brightness), chroma (saturation), or both. In only one of the studies reviewed by Beebe-Center (Dorcus, 1926) was hue varied independently. Beebe-Center (p. 306) organizes the pertinent evidence from Dorcus's study and regards it as supporting an age shift

from warm to cool preference. The evidence, however, is actually very slight; with saturation controlled, it depends entirely upon a difference between grade-school children and college students--hardly comparable samples of their age groups.

Our intention here is to follow the example of Guilford and Smith (1959), Granger (1955), and Helson and Lansford (1967) in their studies of adult color preference, by investigating with children the separate influence of hue, chroma, and value. We deal most fully with hue and chroma but provide some exploration of value also. Preferences are analyzed separately for the two sexes.

METHOD AND SUBJECTS

Our general method of measuring color preference was to show pairs of color patches to classroom groups and have each child record on a mimeographed form whether he liked better the one on the left or the one on the right. Each group was shown the same 35 pairs in a uniform order planned to separate pairs of similar character and two occurrences of an identical color. The 35 pairs fall logically into sets which will be described under Results and Discussion; each pair presented two colors differing on either one or two of the three basic dimensions of color. Nothing was said to the subjects about the nature of these differences. Each pair was shown for about 10 seconds.

The color patches were Munsell papers 5 1/2" x 8 1/2". Each was surrounded by a cardboard matting of flat black, 8 1/2" x 11". On an easel, placed in the classroom to be illuminated by sunlight as much as possible while avoiding glare, rested a 2' x 3' wooden background painted flat black, with L and R printed in chalk on the left and right sides of the top to accommodate children in the lower grades. The two cards to be shown at a time were placed in the

middle of this background approximately 6" apart. (Black surrounds were chosen rather than grey in order that all patches would differ greatly from the background in value.) The experimenter, a senior in Yale College, was introduced by the classroom teacher, and spoke about as follows:

This morning I am going to give you a color-preference test. You are going to choose between two colors and decide which one you like better.

The way I am going to do this is that I am going to place two colors [doing so] on this black board. You are to decide which color you like better [adding in lower grades: which color is more attractive to you, appeals to you more, looks prettier, the color which you like better]. If the color you like better is on the left underneath the L, then you circle the L on the answer sheet. If the color you like better is on the right underneath the R, then you circle the R on the answer sheet.

Each pair of colors has a number and I will say the number as we go along. This is Number One. Decide which color you prefer and circle the appropriate letter.

Are there any questions as to what we are going to do?

Subjects were more than 1100 students in the public-school system of a town which may be characterized as a small middle-income suburb within a metropolitan area of about 250,000. All students in two primary schools (grades 1-6), and five classes at each grade level from the junior- and senior-high schools (grades 7-9 and 10-12, respectively) were given the test. The junior-high school divides its students into classes according to scholastic ability and performance, and an attempt was made to sample over all levels. In the high school, students had to be obtained through a teaching department. The

science department very generously assisted us, and we were able to test 11th- and 12th-grade classes in physics and chemistry and 10th-grade classes in biology.

Incomplete protocols were discarded before beginning to analyze the data. Of 81 first-graders, 37% turned in incomplete papers, as did 16% of the 86 second-graders. That so many in these two grades were clearly unable to follow instructions may make us wonder whether some of their fellows only seemed to follow instructions and gave answers not truly indicative of preferences. In the higher grades, there is no such reason to doubt the validity of responses, as incomplete papers were very few.

RESULTS AND DISCUSSION

A. Complementary hues and preference

1. Preference when hue alone varies

Subjects were presented with eight pairs in which the two colors differed only in hue, being complementaries of the same value and chroma. Each of these pairs may be visualized in Figure 1, where each small circle represents one of our stimuli projected on a cross-section of the Munsell color solid; a hue pair is made up of the two points on a single diameter which are equally distant from the center, one on each side of the center. On each of the four diameters used, there are two pairs of this kind. One pair contrasts complementary hues at high chroma (chroma 8 if possible--if not, then chroma 6); the other contrasts complementary hues at low chroma (4 points lower on the Munsell scale than for the first pair).

The results for these 8 pairs are presented separately in Figure 2, which consists of a separate graph for each pair, each graph showing a line for boys and a line for girls. Each line connects the 12 points representing, grade by grade, the proportion of children who preferred the first-named of the

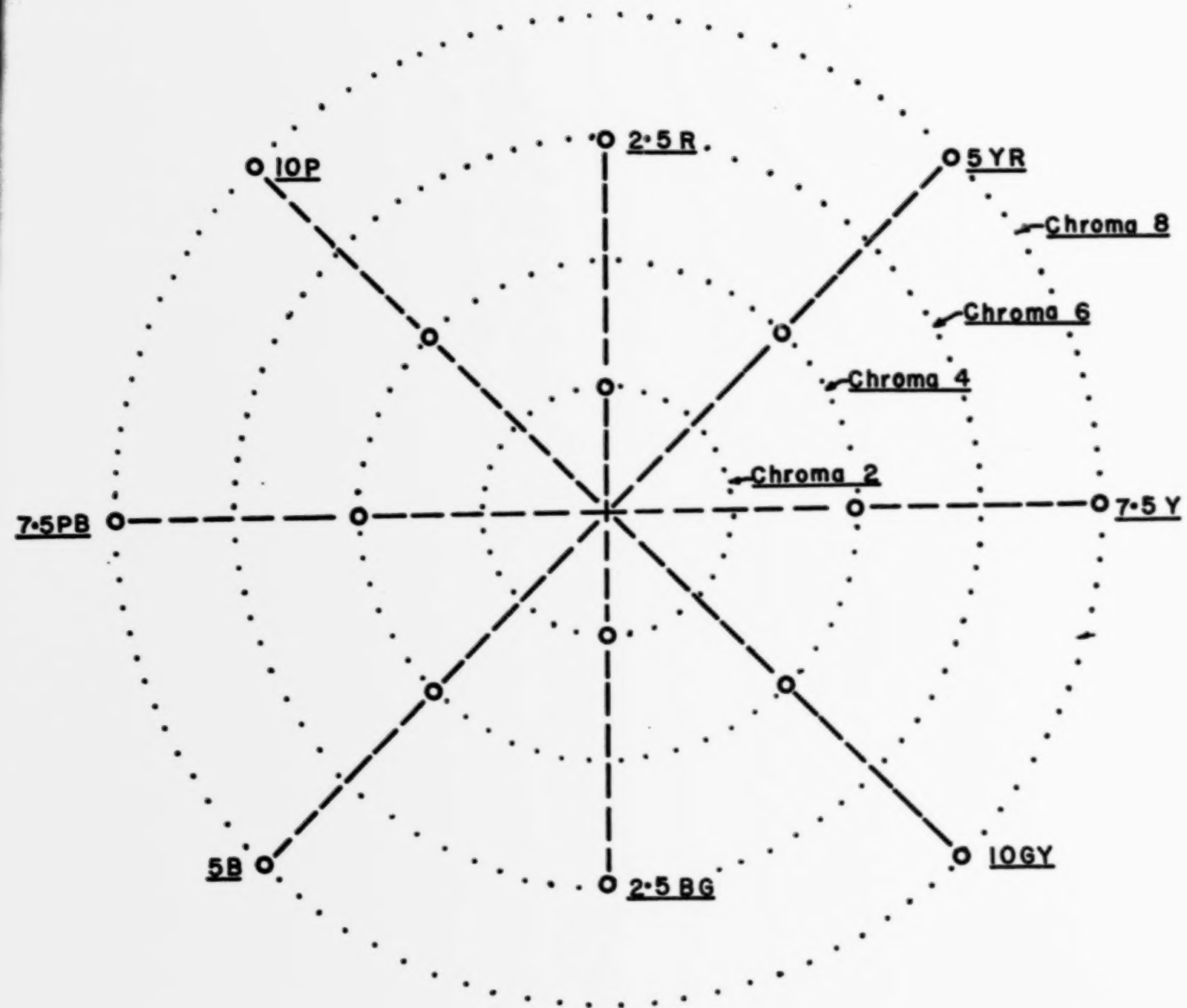


Figure 1

Hue-Chroma diagram.

(Value dimension is perpendicular to this plane.)

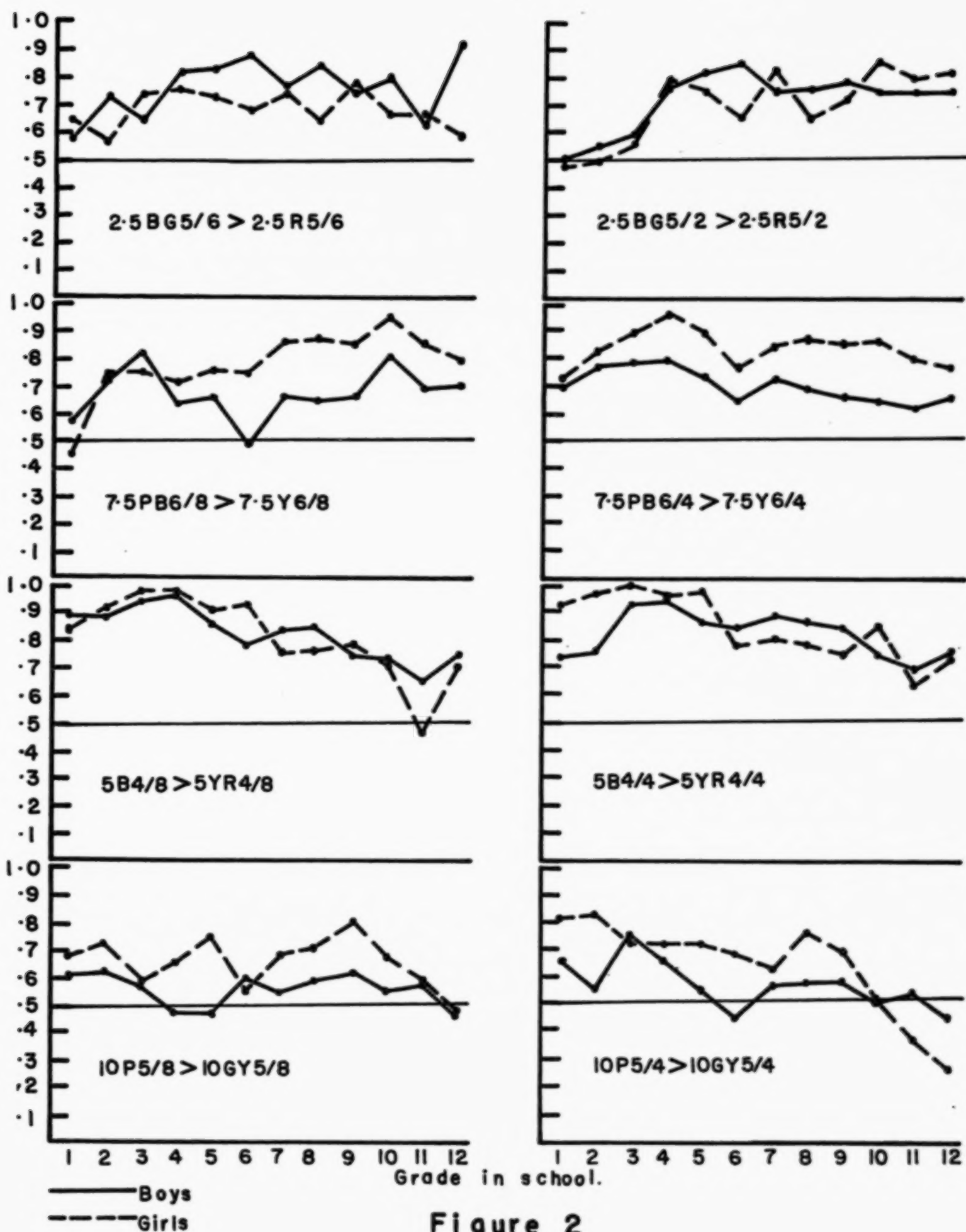


Figure 2

Preference for first-named hue over its complementary.
(In upper 6 charts, preference for cooler hue.)

two colors making up the pair. (The colors are named by the Munsell color system. The letters are initials of familiar hue-names--B for blue, YR for yellow-red, etc.--preceded by a number which specifies the hue more precisely. Following the letter are two numbers separated by a slash. The number before the slash is a measure of value, i. e., lightness; the number after the slash is a measure of chroma, i. e., saturation or purity.) For every pair and for each sex there is clearly a tendency toward consistent preference; none of the lines waver around the .5 level. For most pairs, too, there is definite change with age in the strength of preference; in the pairs represented at the bottom there is even a suggestion that in the very oldest groups the change is great enough to reverse the direction of preference.

Do these results justify a generalization about the kinds of hue children prefer and about the way their preferences change with age? In seeking to answer this question, we will first restrict ourselves to the results for grades 4 through 12; results for the earliest grades may be of uncertain meaning, as already indicated, and we will turn to them later.

In generalizing about hue preferences we find it useful, as did Beebe-Center, to speak about a dimension of cool-to-warm. We are following artistic convention in referring to red and orange-red hues as warm, and blue and blue-green as cool, but find it convenient to go beyond this convention in thinking of these extremes as defining a dimension. From the warm pole of the hue circle, somewhere around 5YR in Munsell terminology, it is possible to shift hue toward the cool pole by moving around the circumference in either direction. Three of the four diameters we used move very decisively from a warmer hue at one end to a cooler hue on the other. Only one of the four diameters--that from 10GY to 10P--is about equidistant on each side from cool and warm pole.

Six of our pairs, then (the uppermost 6 in Figure 2), may be said to

contrast a cooler with a warmer hue, and the cooler hue has always been placed first in the labeling. Thus preference for the cooler hue is indicated by results above .5 in these graphs. The results indicate a strong general preference for the cooler hue over its warmer complementary; of the 144 points entered on these graphs, giving the proportion of an age-sex group preferring the cooler member of a particular pair, only 4 points fall below 0.5 and they are only slightly below.

But what of the supposed preference of young children for warm hues? Even if not present absolutely--i. e., the majority of young children preferring the warmer hues--it might be present relatively, the younger children showing a smaller majority than the older in preferring the cooler hue. Such a finding could be exhibited in a generally positive slope of the age functions for these 6 pairs. It clearly is not; for only two of the pairs is there a definite positive slope, while for three there is a decided negative slope for at least one sex. The particular pairs of complementaries we used show large and sometimes very regular preference changes with age, but these changes are not uniformly related to the cool-warm dimension. Their relation to it may even be reversed by a change in chroma, as may be seen by comparing results of the purple-blue vs. yellow contrast at chroma 8 and chroma 4; for the highly saturated colors, age brings increasing preference for the purple-blue, while for the less saturated colors it has an opposite effect. Because of this clear evidence of great diversity among the pairs, a much larger sample of pairs is needed to determine whether there is a general tendency with increasing age toward increasing preference for cool or for warm hues.

If we now consider the three lowest grades, we find that they often deviate from the general trend of the older children, and that they exhibit a consistency of their own. In every one of the 12 curves relevant to the cool-

warm dimension, the 3rd grade shows more marked preference for the cooler hue than does the 1st grade, and the 2nd usually falls between. Do these findings confirm the common belief in a warm preference in early childhood and indicate that the shift to cool preference occurs very early? Perhaps further research may give them this meaning. At present, however, they seem more plausibly explained as an artifact resulting from the same circumstances that made for many incomplete papers in the 1st and 2nd grades. Young children are the ones most likely to respond irrelevantly, marking the answer sheet in ways having no dependence on the particular stimuli exhibited. The larger the proportion of children responding irrelevantly, the closer the results should approach an equal choice of the two alternatives. The consistent performance of the lower grades shown here is all of this pattern, and most of the curves to be presented or summarized later also show this pattern. (The pairs represented at the bottom of Figure 2, however, are exceptions.)

For hue, results pertaining to sex differences are also indecisive. Two pairs (both contrasting purple-blue with yellow) show a clear consistency, girls preferring the cooler hue more often than do boys. For one of the remaining pairs there is a slight suggestion of an opposite sex difference, and for the other three pairs contrasting cooler with warmer, the results vary from grade to grade.

The 7th and 8th hue comparisons, which fall on a diameter approximately perpendicular to the cool-warm axis, are not likely to be relevant to the cool-warm dimension. Here again, though, at the bottom of Figure 2 we see marked differences in age function according to chroma level. These results also confirm the ones on coolness vs. warmth in suggesting sizable sex differences in response to particular hue contrasts.

2. Hue preference when chroma also varies

A set of 8 pairs presented complementary colors which also differed greatly in chroma. The patches used were the identical ones used in the set already described; here, each patch was paired with that complementary shown in Figure 1 which differed from it in chroma. The results confirm those obtained with the pure hue contrasts; we will not present them, but can briefly summarize them as follows:

- a. A general preference for the cooler hue remains evident despite variation in chroma, although it is less consistent than for the pure hue contrasts.
- b. Striking sex differences and general age trends are found for particular pairs, but they are not consistently related to the cool-warm dimension.
- c. For 7 of the 8 pairs, the three lowest grades show, with lower age, an approach toward the 0.5 point. For the other pair, the curve is practically level for these grades.

3. Hue preference when value also varies

Only 2 pairs simultaneously varied hue and value. They are not relevant to the cool-warm dimension, having been chosen from the P-GY diameter approximately perpendicular to that axis, and the results will hence not be reported here.

B. Chroma and preference

1. Preference when chroma alone varies

The 8 chroma contrasts may also be visualized by reference to Figure 1. On each of the 8 radii shown in this color wheel there are two patches differing only in chroma; each of these 8 pairs was presented to the subjects for preferential choice. The results are given in Figure 3. The lowest three grades show the familiar trend away from 0.5, suggesting again that some of the younger children were not responding to the stimuli. In considering the rest of

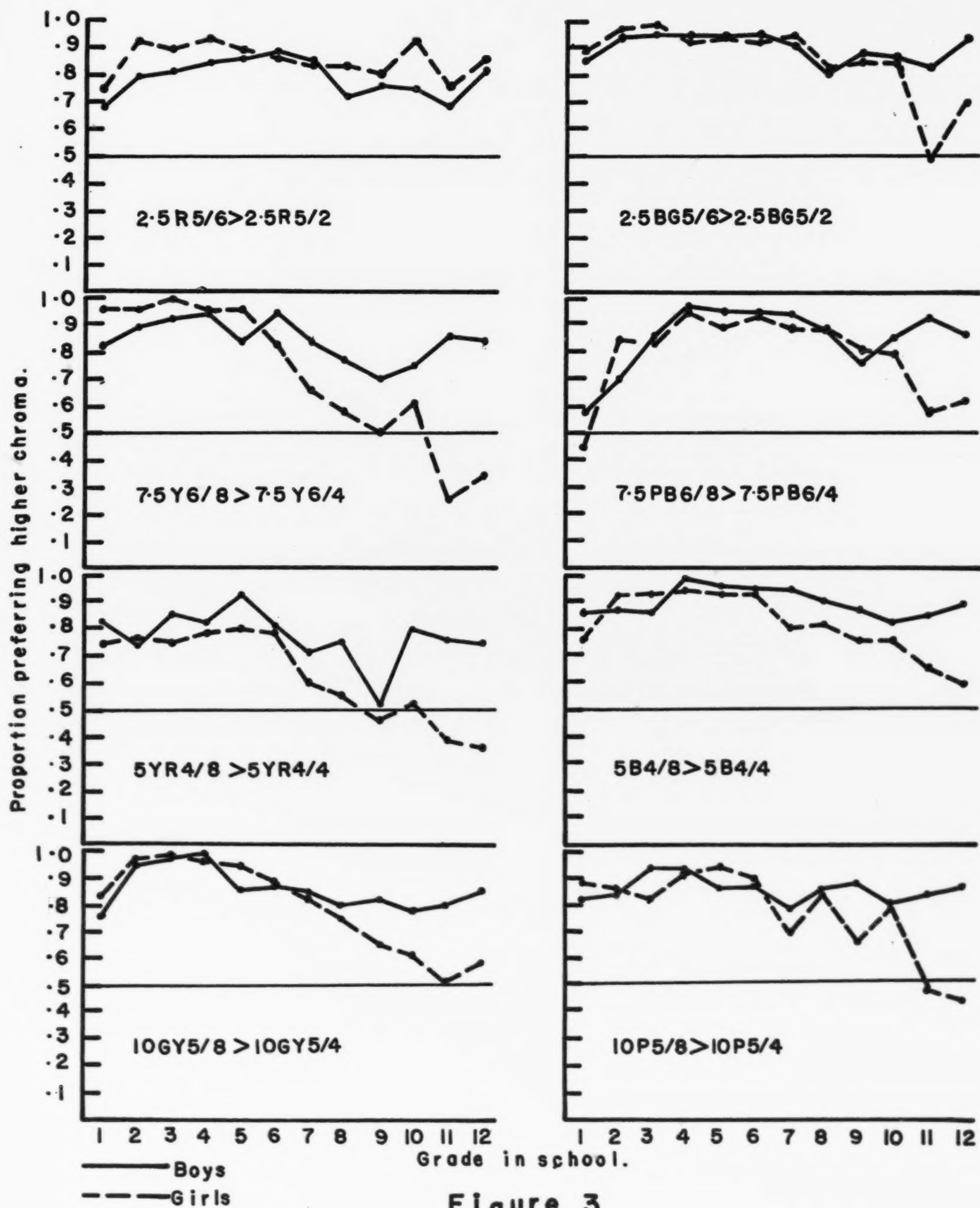


Figure 3

Preference for higher chroma.

the results, therefore, we shall attend principally to the upper nine grades.

First, there is a general preference for colors of higher chroma. On these graphs 192 points are entered (considering here all 12 grades); only 9 of these fall below 0.5, and few fall much below. With such chroma differences as we have used, preference for the more saturated color is very marked. (The differences in chroma represented in these pairs are very conspicuous, though by no means as large as they could be.)

Second, there is a thoroughly consistent age trend. For each of the eight pairs, and separately for boys and for girls, the results for grades 4 through 12 show a clear downward trend. Generally the change is from over 90% preference for higher chroma among 4th-graders down to about 80% in boys and much lower in girls in the high-school years.

Third, as just implied, there is a sizable and rather consistent sex difference in the age trend. In the 4th grade, boys and girls are very similar in chroma preferences. By the 12th grade there is a large difference; in some pairs fewer than half of the girls now choose the higher chroma, whereas this choice is never made by fewer than 75% of the 12th-grade boys. Of the 8 pairs, only the red one provides an exception to this generalization; in high school as in grade school, girls seem even more frequently than boys to prefer high saturation in a red.

2. Chroma preference when hue also varies

The pairs relevant here are the same pairs reported on with respect to hue preference when chroma also varies. The effect of hue on preference is so large as to obscure somewhat the findings reported for choices when chroma alone varies. But the findings are still there, confirming all the tendencies found for variations in chroma alone.

3. Chroma preference when value also varies

Only 2 pairs presented this double contrast. They, like the 2 pairs varying hue and value, were included to observe resolution of conflict, and the results will not be considered here.

C. Value and preference

As mentioned earlier, value was explored only sparsely. Six pairs were presented in which the two patches differed in value alone. They were chosen for suitability to observing resolution of conflict, an aspect of the research to be considered later in this paper. Four of the pairs had as one member one of the patches of value 5 shown on the LOGY-10P diameter in Figure 1, and as the other member the corresponding patch from the parallel diameter at value 7. The other two pairs had as one member one of the two patches of value 4 shown on the 5YR radius in Figure 1 and as the other member the corresponding patch from the parallel radius at value 6.

The results obtained with these 6 value contrasts are pictured in Figure 4. The most dependable finding is of a difference between the sexes; girls choose the higher value in each pair more often than do boys in each of the 12 grades, with only 6 exceptions out of the 72 comparisons represented in the graphs. This finding is consistent enough so that we may reasonably guess it is likely to hold true for other hues and chroma, though only direct test will prove whether the guess is correct. The results also indicate that some pairs show definite age trends, more marked in girls than in boys. The only strong consistency in age trend across pairs is a trend for girls to decrease their preference for high value during the high-school years. With respect to general preference for higher or lower value, there is a suggestion of general preference for higher value (especially in girls), but inter-pair variation is great

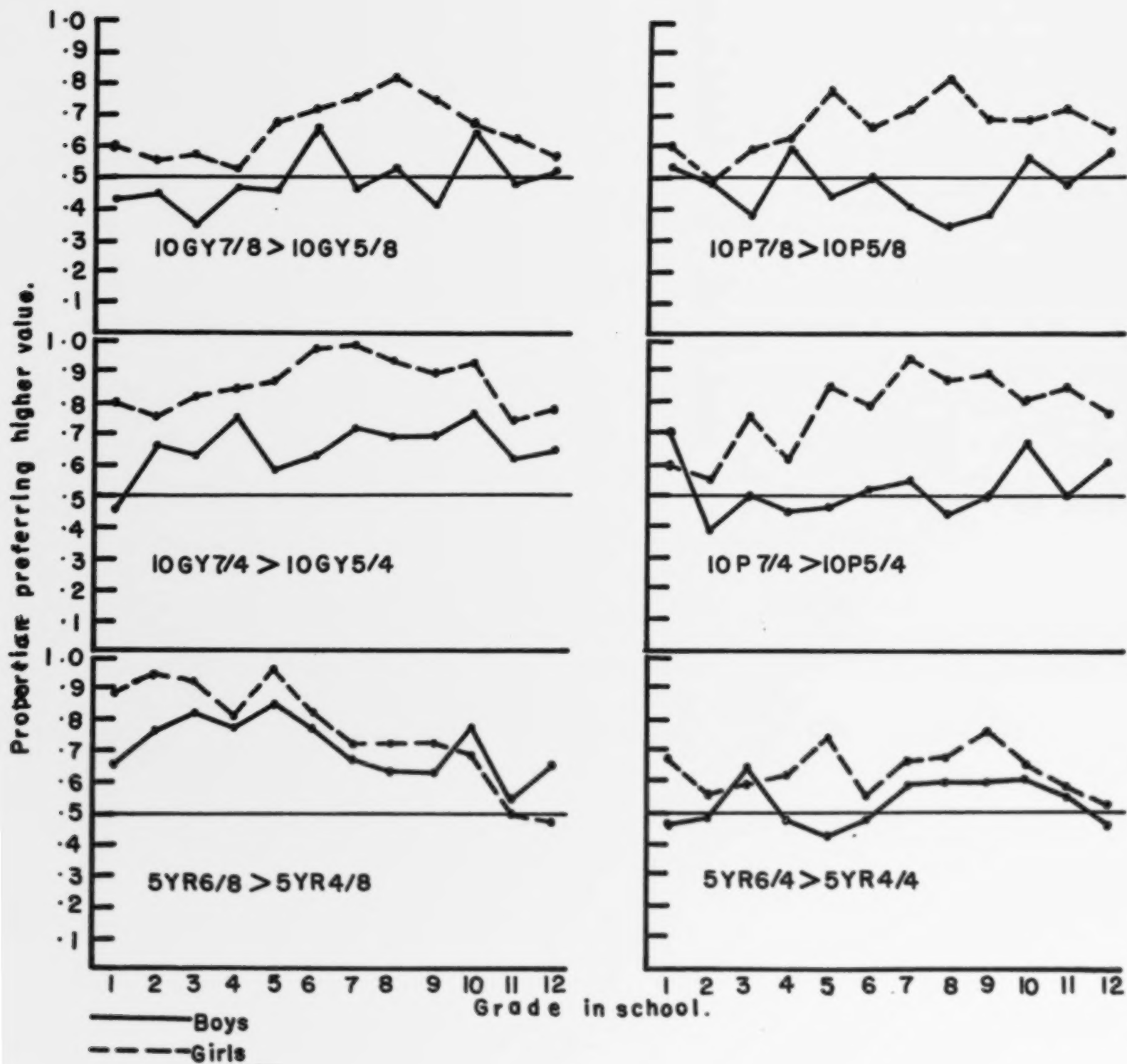


Figure 4

Preference for higher value.

enough to suggest caution in predicting value preferences with other hues and chroma.

D. Relative consistency of hue and chroma preferences

Choices made on the hue contrasts and on the chroma contrasts can be looked at to see whether the relative consistency of the two influences changes with age. For this and the analysis in the following section, only a portion of the children contribute data relevant to each question; so that to obtain samples of adequate size we need to combine groups. We combined on the basis of age rather than class, and divided subjects into those aged 7 through 12 and those aged 13 through 18. (None were older than 18, and those under 7 were omitted as a device for eliminating most of the children who did not follow instructions carefully enough.)

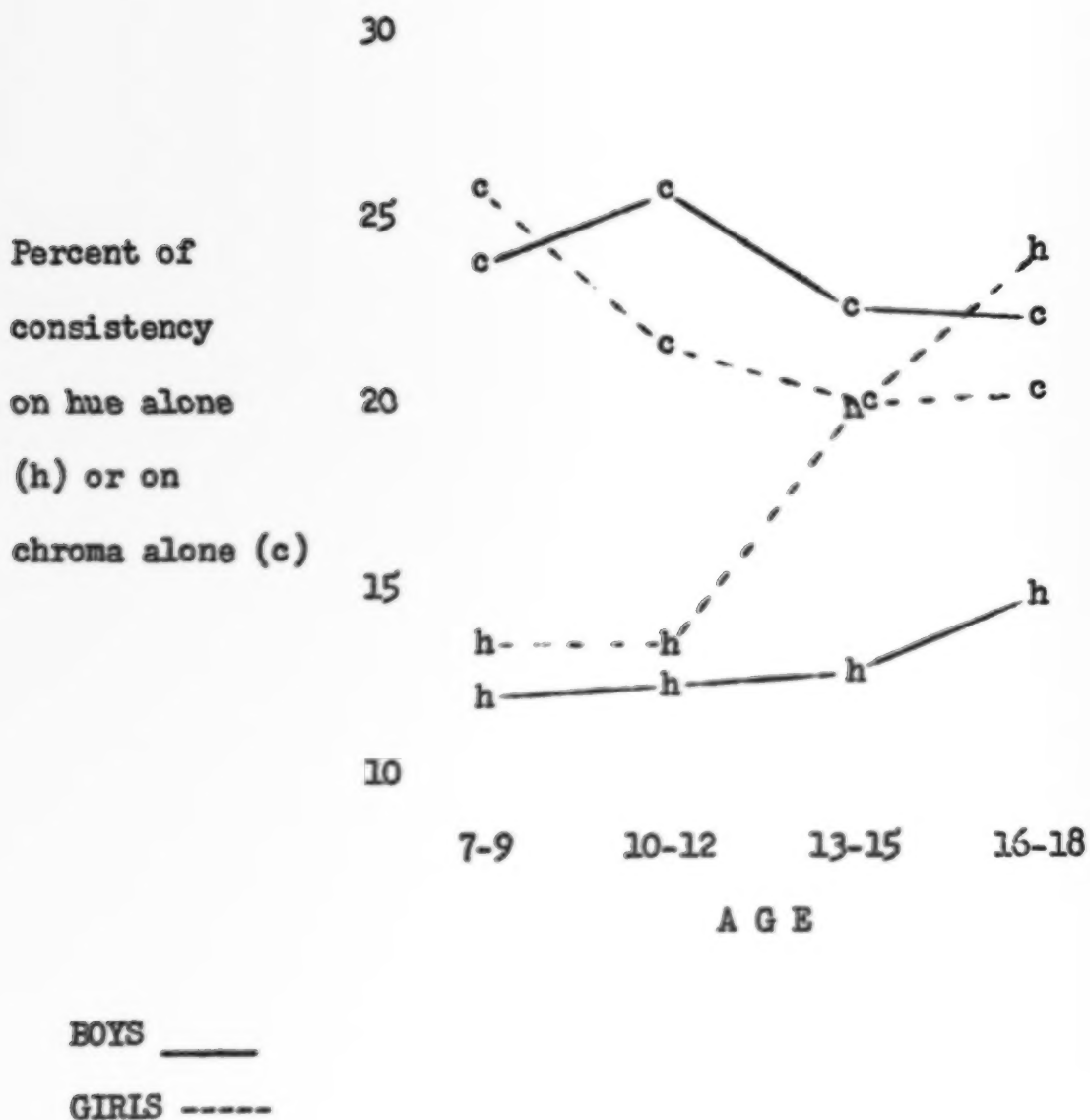
We have seen that on each of 4 diameters on the color wheel a child was presented with two simple hue contrasts and with two simple chroma contrasts. Would he respond consistently to the two hue contrasts, preferring the same hue to its complementary regardless of what chroma level characterized the pair? And would a child respond alike to two chroma contrasts, choosing low chroma (or high) consistently, regardless of whether the two patches were both of the one hue or were both of its complementary? In a slight majority of instances (55%), both the two choices a child made when faced with hue contrast and the two choices he made when faced with chroma contrast were internally consistent. In a small minority of instances (7%), the choices a child made along one diameter were consistent on neither dimension. The remaining instances, where the child's choices were consistent on one dimension and not on the other, are the ones we must consider in order to compare consistency on the two dimensions.

Figure 5 displays the outcome. A clear trend seems to be present in both sexes, though much larger in girls. With increasing age, hue contrasts

Figure 5

Consistency of choice on dimensions of hue and chroma, for four age groups.

(Each entry shows the percentage of instances in which, on a single dimension of the color wheel, a child was consistent in hue choices alone or was consistent in chroma choices alone.)



are responded to more consistently and chroma contrasts less so.

To assess the statistical significance of the difference, we need to get down to the individual. We need consider a person only if on at least one diameter of the color wheel he was consistent on hue alone or on chroma alone. What we are interested in is, How many more (or fewer) times is he consistent on hue alone than on chroma alone? This score could vary from +4 to -4, and a few individuals reached even to these extremes. We can then ask whether older and younger children differ significantly in mean score. The older children have a higher mean by 0.36 of a unit on this scale, and for boys and girls together this age difference is highly significant, with a t of 3.85. For boys the difference is 0.20 and for girls 0.52; the former is not separately significant, while the latter is (t is 1.53 and 3.94).

Thus we have a significant change with age in the relative consistency of hue and chroma preferences. The age effect does not appear with perfect consistency on every one of the four diameters we sampled from the color wheel. To test satisfactorily whether there are real differences in this effect among the diameters would require another investigation. That such differences occur in our data, whether sampling error or real effect, means that we can draw no definite conclusion about relative consistency of hue and value preference, for this was tested on only a single diameter. On this one diameter there was a slight and irregular trend with age toward greater consistency of value choices and lesser consistency of hue choices.

These findings on consistency changes may be regarded as not altogether novel in relation to the findings we have reported earlier in the paper. Where choices are made almost unanimously, and in the same direction for the two hue contrasts on a single diameter, or for the two chroma contrasts, consistency within the individual must necessarily characterize almost everyone. The more

that choices fall away from unanimity, the greater is the possibility for inconsistency within the individual as well as between individuals. The changes in consistency we have just considered are of such a pattern that they might express only this fact. They do not necessarily follow from it. We are dealing here with consistency of choice regardless of which choice is made. People might be evenly divided on whether they preferred high or low chroma, yet each person's preferences could be internally perfectly consistent. To be sure whether the changes in consistency are at least operationally a separate phenomenon, we need to base them on data independent of the general trends established earlier. For this reason our pairs included certain ones intended to put hue preference-tendencies and chroma preference-tendencies in competition with each other. We will now look at the outcome of such competition.

E. Resolution of conflict between hue and chroma preferences

For any single diameter of the color wheel, we will consider here only those individuals (a slight majority) who make perfectly consistent choices on the simple contrasts we have been discussing. The vertical diameter of Figure 1 will serve as an example. At some time during our session the individual sees two red patches which differ only in chroma, and chooses (let us say) the one of lower chroma; at another time he sees two blue-green patches and again chooses the one of lower chroma. He also sees a high-chroma red and an equally saturated blue-green, and chooses (let us say) the red; then, seeing a rather low-chroma red and blue-green, he again chooses the red. This is what we mean by saying he makes consistent choices on the simple contrasts; in this case the choices consistently express a preference for red and a preference for low chroma. The individual also sees, however, two more complicated pairs posing contrasts simultaneously on the dimensions of hue and chroma; in one, a low-chroma red is paired with a high-chroma blue-green, and in the other a high-

chroma red is paired with a low-chroma blue-green. For the first of these two pairs he can remain consistent with both his general tendencies by choosing the low-chroma red; the second permits him to remain consistent with only one. He need not be consistent at all, of course; to choose, in the first of these pairs, the high-chroma blue-green would at once deny both of the tendencies he had expressed in response to the simple pairs. If in one of these complex pairs he is consistent with both tendencies expressed on the simple pairs--as our subjects were in more than 90% of the instances--but a second complex pair forces him to choose between them, to which of the two tendencies will he then be faithful? (For clarity of presentation we have spoken here as though the subject first expresses tendencies on simple contrasts and then decides subsequently what to do about them on complex contrasts. In fact, the various contrasts on different diameters were presented fairly swiftly in a jumbled order, as we indicated earlier, and few subjects are likely to have formulated any clear idea about how the various pairs related to one another; though we made no inquiry, we suspect that almost no one was aware that we were measuring general tendencies and then putting them in competition to observe the outcome.)

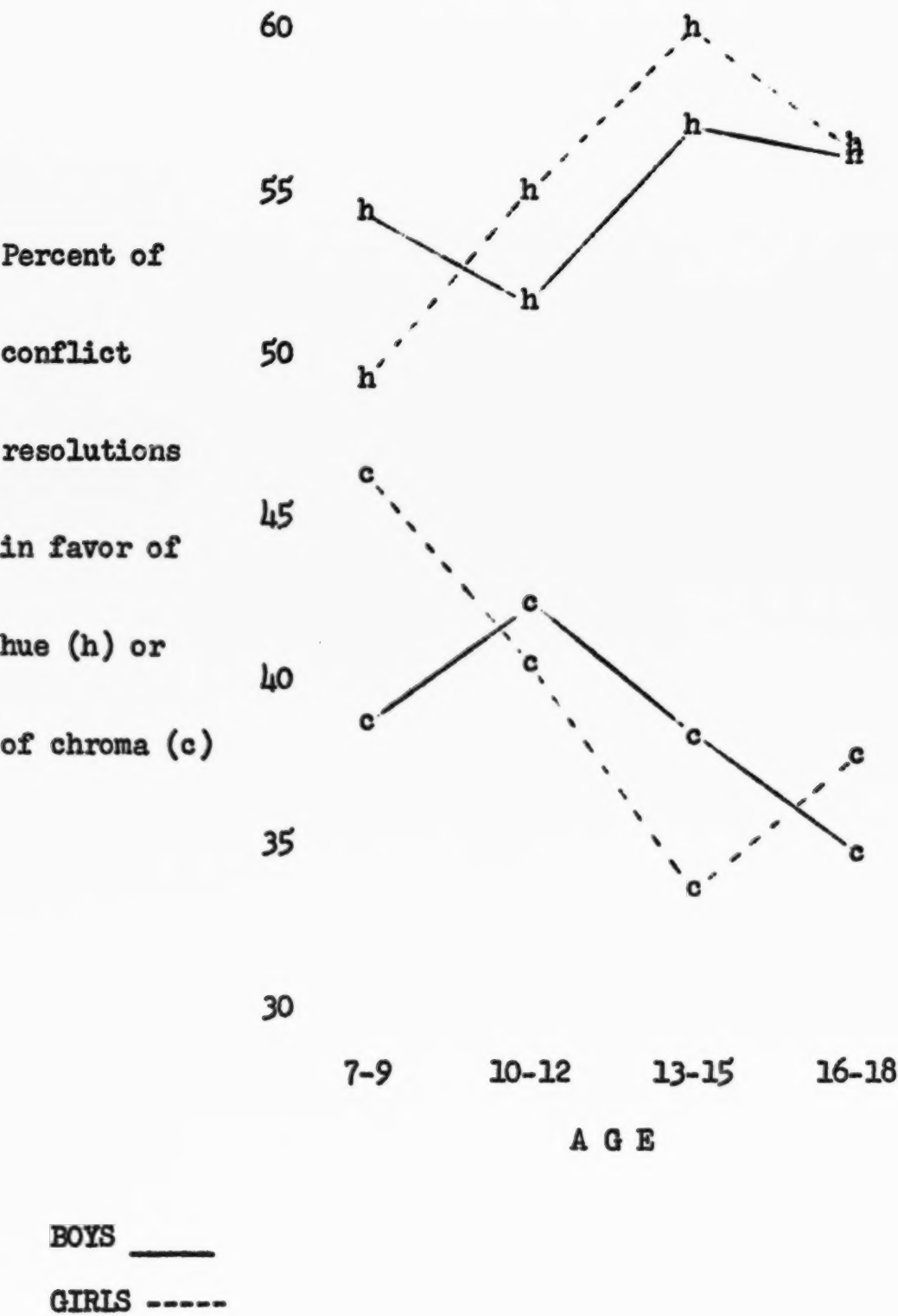
Figure 6 gives the answer. We note first that hue appears to be the more important influence at all ages. We must not take this very seriously, remembering that the hue difference we employed is the greatest possible, whereas the difference in chroma is always of just 4 points--never more than half the range available on Munsell papers and for some hues only a quarter of the range. What is important is the relative influence of age on the two variables. With increasing age, conflict is more often resolved in favor of hue and less often in favor of chroma. Again, the age change seems to be greater for girls than for boys.

To test the significance of the trends presented in Figure 6, we again

Figure 6

Resolution of conflict in favor of hue or chroma, for four age groups.

(Each entry shows, for instances where on one diameter of the color wheel a child was consistent both on hue choices and on chroma choices, the percent of conflict resolutions in favor of each dimension.)



need to score individuals on strength of the difference. In response to the simple pairs people need to be, on at least one diameter, consistent on both dimensions in order for their further responses to be relevant, and for only such diameters will their responses be relevant. For each such person we count the number of diameters on which he resolved a conflict in favor of hue and the number on which he resolved a conflict in favor of chroma. The difference between these, the extent to which his resolutions in favor of hue exceed his resolutions in favor of chroma, is the measure we need. The difference between age groups on this measure may then be tested for significance by calculating t for the difference between means. Subjects aged 13 through 18 average 0.24 higher than younger subjects, and this difference for the two sexes together is significant at the 2% level ($t = 2.49$). For boys alone, the difference is 0.20, for girls 0.28, and only the second of these is significant by itself (t is 1.46 and 1.96).

Again, this significant over-all finding does not appear on every separate diameter sampled from the color wheel. It appears with great consistency on the R-BG and Y-PB diameters; and on the GY-P diameter only a very deviant figure for the oldest girls departs from the same trend. But on the fourth diameter, YR-B, the results are entirely contrary for both boys and girls. Again, too, this discrepancy among diameters prevents us from placing any confidence in a similar comparison of hue with value, for this is possible only on one diameter, GY-P. For that one diameter, the results are extremely regular, exhibiting a steady decrease with age in resolution of conflict toward value choices and corresponding increase in resolution of conflict toward hue choices.

DISCUSSION

A. General preference tendencies

Hue. Our finding that when chroma and value are held constant, cooler colors are preferred by both sexes and at all ages (with some possible exceptions only for the oldest girls) extends to the years of childhood and adolescence a generalization already well established for adults in England and the United States. The three major studies all provide clear evidence that adults generally prefer cool colors, although these terms are not always used, and sometimes other hue preferences also appear (Granger, 1955; Guilford & Smith, 1959; Helson & Lansford, 1967). Despite overwhelming support for the fact, as yet no studies seem to be particularly directed at its interpretation. Our study makes one contribution here. We find that a uniformly administered technique for measuring hue preference produces, for some hue choices, very different results at different ages, altering greatly the size of the majority preference for the cooler color or, at least in hue choices not relevant to warmth vs. coolness, possibly reversing the direction of preference. This finding of large age variation in preference disposes of the extreme view that hue preferences are entirely determined by inherited biological determinants uniform throughout mankind. Whether simple uniform biological determinants are, however, important, and what else is, remains to be discovered in the future.

Chroma. Our finding that higher rather than lower chroma is preferred extends to childhood and adolescence another generalization previously established for adults. One of the previous studies reports an upturn of the curve at lower chroma, and another reports a downturn at higher chroma, than any we used. But over the range of chromas we employed, all three major studies are in agreement that preference is a positive function of chroma. Our study resembles those by Granger and by Guilford and Smith in displaying the color

patch on a background of zero chroma--i. e., gray or black. These studies alone might leave open the possibility that what is liked is chroma contrast rather than high chroma itself. Helson and Lansford, however, report the same results even though they used backgrounds of varied chroma. Again, no research as yet seems to have been mainly directed at testing hypotheses about the reasons for this general preference. Our finding marked age difference does, as in the case of hue, decisively disprove the simplest view of the preference as solely determined by a uniform inherited disposition.

Value. Our findings contain only a suggestion of a general tendency to prefer higher value. The uncertainty here is a good fit with the fact that the major studies of adult preferences differ radically from one another on the effects of value. No broadly established generalization can be made.

B. Sex differences

Granger (1955) felt that his findings justified the view that color preferences do not differ appreciably between the sexes. So far as American children and adolescents are concerned, our findings decisively refute this view. At the same time, our findings do indicate a general similarity of the sexes in color preferences, and this point seems to be all that is really justified by Granger's data. The other major studies also agree in finding important sex differences along with general similarity. The only specific comparison possible is with Helson and Lansford's finding of relatively greater female preference for warm colors and male preference for cool colors. We do not find any consistency in this direction. This difference between the sexes may develop at a later age than 17, or the discrepancy between the studies may be attributable to sampling error associated either with Helson and Lansford's having so small a sample of subjects or with our having an even smaller sample of stimuli relevant to the warm-cool dimension.

C. Simplicity of relationship

Each dimension of color experience can be considered a determinant of preference. Does each dimension influence preference according to a simple law which remains the same regardless of the levels reached by the other dimensions? One extreme position is that of Granger (1955), who argues for this simple uniformity. At the other extreme is the "clinical" view of some artists that each color has emotional meaning or evokes like or dislike in a way that is largely sui generis. Intermediate is the position of Guilford and Smith (1959), who in effect argue for a rational model of color preference, but one recognizing interaction of dimensions--that is, expecting changes in effect of one variable as other variables change. Our results have some bearing on that issue. In finding that the influence of value (for example) on preference varies with age in a very different way at different hues and chroma, our study seems to indicate that for particular ages too a single set of simple laws cannot apply throughout. Thus we add substantially to the evidence against a perfectly simple system. But the extreme view of specificity of effect--that the effects of colors are too specific to be usefully related to any abstract generalization--is also not compatible with our results, nor is it compatible with any previous results we know of. Perhaps some new evidence on this point is provided by our pairs composed of simultaneous contrasts in hue and chroma; all the simple effects of hue and of chroma could still be seen in this more complex situation. On the whole, our results seem most compatible with the complex-system view of Guilford and Smith, though we have not sampled the color solid thickly enough to test their or any other particular system.

D. Change with age

On age change, no comparison with earlier studies of the three color dimensions is possible. We can compare our results with the common opinion that as children grow up they shift from preferring warm hues toward preferring cool

ones; we find no evidence for any general change in this direction. We do, however, find clear evidence of a general weakening with increased age of a childhood preference for high chroma. We also have significant evidence from two other sources that chroma decreases with age in its importance, relative to hue, as a determinant of color preference: chroma choices become relatively less consistent with increasing age, and hue choices more consistent; and conflicts between chroma and hue preferences tend to be increasingly resolved in favor of hue.

The diminishing importance of chroma at a higher age, and the greater importance of hue, may be viewed as an instance or consequence of increasing differentiation. A simple way of putting the hypothesis would be this: In responding to chroma we are being influenced by what diverse colors share--their common colorfulness; the only distinction we make is in the intensity of this general characteristic. In responding to hue, we are being influenced by what is most distinctive about diverse colors, their unique quality. Doing so seems to involve, therefore, a more finely differentiated response to the environment. (Koffka [1925, 264-284], in discussing the development of color perception in infancy and early childhood, makes a similar point about the role of hue and chroma as bases for perceiving differences.)

Another way of stating the hypothesis would call attention to the kind of changes involved in varying chroma as against varying hue. Chroma variation can be satisfactorily represented as a single dimension from complete colorlessness to maximum intensity or purity, with the same kind of change occurring in all sections of the dimension. Hue cannot be represented so simply. It is a variable which goes around in a circle to return upon itself, and it obviously cannot be varying in the same way everywhere. The recent application to hue of direct scaling (Indow & Stevens, 1966) provides a specific illustration of the

fact that change in this variable differs in character from one section to another. Preference based on hue, then, may be said to imply more differentiated response to the environment than does preference based on chroma.

Our results suggest that the widespread belief that children prefer warm hues may have only an indirect validity. We find that children, like adults, when choosing between colors differing only in hue, generally prefer the cooler. In normal life environments, however, the several color attributes vary simultaneously and are intercorrelated; the most highly saturated colors are warm in hue. That is, most warm hues can have higher chroma than can most cool hues. In the color samples sold by the Munsell Color Company, for example, 5YR is available at chroma as high as 12, whereas 5B is available only at chroma up to 8. If many warm colors in the child's environment appeal greatly to him, then, they may appeal because they are much more saturated than the cool colors he sees, not because they are warm--in short, because of their chroma, not their hue. Our hypothesis thus accounts for some of the differences between adults and children in color preferences by viewing them as a consequence of the greater differentiation of cognitive functions in the adult.

This hypothesis might be opposed on the grounds that the findings most directly supporting it--age changes in consistency and in conflict resolution--are small effects. Against this argument we might point out that the increase in cognitive differentiation from age 7 to adulthood is probably rather small on an absolute scale, too, however great its consequences.

This hypothesis has some testable consequences:

1. The hypothesis seems equally applicable to the relative importance of value and hue, and should there predict similar changes with age. Our data offer only limited opportunities for testing this prediction; on only one diameter of the color wheel did we compare value and hue as bases for preference

and put them into conflict with each other. On this one diameter, as we have already indicated, one of the predicted changes with age was strikingly confirmed and another was not.

2. The hypothesis relates increasing predominance of hue over chroma and value to cognitive growth which should characterize persons in any normal environment. Whereas other aspects of color preference may differ a great deal from one society to another, then, we would expect that this feature--the increase with age in relative influence of hue--might be found everywhere present in some form.

3. Predictions about preferences in other sensory modalities seem possible even though the parallel to vision is far from perfect. The quality of a taste or smell compared with its strength, for example, or the timbre of a tone compared with its loudness or pitch--those qualities might, like hue, be expected to gain importance with a person's age as influences on his preferences.

4. Our hypothesis suggests that similar changes will be found at earlier ages than we have explored, if younger children are tested for color preference by methods suited to them (as our methods clearly are not). If we are right in surmising that the changes in cognitive differentiation between birth and the age of 7 are much greater than those in the following 11 years which we have studied, then we might expect that color-preference changes resulting from increased differentiation would be even more rapid and conspicuous at lower ages. It is encouraging to note that Beebe-Center, in summarizing the poorly controlled studies of infantile preference available to him, concluded, "In the first half-year, preference depends only upon saturation and brilliance" (1932, p. 306).

5. Our hypothesis yields a prediction, too, about variation of color preferences among children of a single age. Mentally retarded children should differ from average children of their age in general cognitive skills in much

the same way as do average children of a younger age. Children of superior intelligence should differ from average children of their own age in the opposite direction. If the relative influence on color preferences of hue as against chroma and value is determined by general level of cognitive development, then it should vary systematically among children of the same age according to their general developmental status.

REFERENCES

- Beebe-Center, J. G. The psychology of pleasantness and unpleasantness. New York: Van Nostrand, 1932.
- Ejersted, A. Warm-cool color preferences as potential personality indicators: preliminary note. Perceptual and Motor Skills, 1960, 10, 31-34.
- Burnham, R. W., Hanes, R. N., & Bartleson, C. J. Color: a guide to basic facts and concepts. New York: Wiley, 1963.
- Dorcus, R. M. Color preferences and color associations. Journal of Genetic Psychology, 1926, 33, 399-434.
- Granger, G. W. An experimental study of colour preferences. Journal of General Psychology, 1955, 52, 3-20.
- Guilford, J. P., & Smith, Patricia C. A system of color-preferences. American Journal of Psychology, 1959, 72, 487-502.
- Helson, H., & Lansford, T. The role of spectral energy of source and background color on the pleasantness of object colors. Illuminating Engineering, 1967 (in press).
- Indow, T., & Stevens, S. S. Scaling of saturation and hue. Perception and Psychophysics, 1966, 1, 253-272.
- Wright, B. The influence of hue, lightness, and saturation on apparent warmth and weight. American Journal of Psychology, 1962, 75, 232-241.
- Koffka, K. The growth of the mind. New York: Harcourt. Brace, 1925.

BaKwele and American Esthetic Evaluations Compared¹

Irvin L. Child
Yale University

Leon Siroto
Chicago Natural History Museum

In matters of value, the hard-won knowledge of cultural relativity contributed by anthropologists to mankind has helped shatter the ethnocentric absolutism of Western thought. But what should replace the earlier belief in absolute universals of value? One possibility is a complete cultural and personal relativity of values, and many people today probably believe they are in tune with the wisdom of anthropologists in taking that position. But clearly this opinion does not prevail among anthropologists. In a discussion of ethical relativity, for example, Kluckhohn (1955) regards both extreme relativism and extreme absolutism as untenable.

The reasons with which Kluckhohn supports his rejection of complete ethical relativism seem in part to be applicable, with appropriate modifications, to esthetic relativism as well. Some anthropologists have made quite explicit their doubt or rejection of complete esthetic relativism. Firth (1951: 161), for example, in an extended discussion which takes account of the facts of cultural variation as seen by the field worker, rejects the extreme position outright in saying, "I believe that there are universal standards of aesthetic quality, just as there are universal standards of technical efficiency." Linton (1954: 166), while taking no explicit position, implies at least a hunch that esthetic relativity is likely to be found untenable:

Whether there is some denominator common to all the expressions of a single type is one of the most important problems confronting the student of esthetics. It is obvious that appreciation of any particular art form is to some degree a result of learning and habituation. Thus, to a European, most African art is repulsive at first contact. It is only after he has become accustomed to the medium that he can appreciate its qualities and derive esthetic satisfaction from it. The real problem is whether behind such diverse objects as a Poro mask, the Venus de Milo, and a Peruvian jar there are common factors of form, dynamic interrelation of parts, harmony of color, and so forth, which may appear in different combinations but are responsible for the esthetic effect. It seems that we have here an area in which modern psychological techniques could be brought to bear on a problem which philosophers have discussed for centuries without coming to agreement.

The appeal for empirical evidence on the question of esthetic relativity has been answered by actual research to an astonishingly small degree. One of the first to use quantitative methods in trans-cultural investigation of es-

thetic values was an anthropologist, Robert Lowie (1921). Aware of psychological research demonstrating some tendency in Western subjects to make and to prefer rectangles shaped approximately in accordance with the golden proportion, Lowie studied the distribution of rectangle shapes in the decorative art of a particular North American Indian tribe, and reported with interest their failure to conform to this proportion. His report is an extremely valuable beginning, for which Lowie claimed no more. That it was not followed up may be traced to the defects of its background in experimental psychology, where the variables then selected for study were probably too simple to get near the heart of esthetic value. More recently, psychological esthetics has shifted attention to more complex variables and to actual works of art as well as to elements used in the arts. In line with this shift, actual works of art (or reproductions of them) have been used in the other two quantitative investigations known to us. Both were concerned with the question of trans-cultural agreement in the order of preference for a set of works of visual art. In one of these studies, Lawlor (1955) compared the preferential ratings of a few West African decorative designs by West Africans and by Europeans. In another, McElroy (1952) compared preference ratings of a variety of visual materials by Australian aborigines and by students at the University of Sydney. In neither of these studies was any evidence found of trans-cultural agreement.

Each of these studies has a serious defect: the people whose judgments were used were not selected for interest in art but rather were representative in this respect of the community as a whole. There seems no reason to believe that esthetic values would be shared by all members of the average community; in a study of college students, Child (1962) learned that preferences according with traditional esthetic values are clearly not those of the average student. In this respect esthetic values may well differ from moral values. Moral values have to do with the evaluation of alternative resolutions of conflicts experienced by everyone, and thus have a necessary relevance to everyone's life. Esthetic values have to do with the evaluation of stimuli with respect to their adequacy for satisfying a kind of interest which most individuals seem to have the option of pursuing or not pursuing; in this sense, they seem to lack any compelling relevance to every individual's life. It is possible that esthetic evaluations may be made by or known to only some people in each society, and yet that agreement will be found between such people in various societies.

Accordingly, we need to ask whether people interested in art within one cultural tradition will have any tendency to make preferential judgments which will agree with those made by people interested in art within another cultural tradition. In the attempt to do this it seems important to use kinds of work which can readily be regarded appreciatively in both societies. Experts in our society are now so used to looking at art from a variety of cultures that they can so regard any visual art. To people in a simpler society who are interested in art, however, art from outside their own cultural tradition might be so novel that they would be unable to consider it appreciatively. One possible approach, therefore, is to obtain preferential reactions to

art works of another society from people in that society interested in those works and then compare their reactions with those given to the same works by art experts in our society. We report here what is, so far as we know, the first study of this sort ever done. It was made possible by a field trip one of us (Siroto) made, devoted primarily to studying the function of masks in BaKwele culture.

The BaKwele are a Bantu-speaking people living in the heavily forested basins of the Dja and Ivindo Rivers in the Republic of the Congo (Brazzaville) and the Gabon in western equatorial Africa. At the time when masks were used, the BaKwele were swidden cultivators who moved their settlements about often and erratically. Numerous patrilines formed large villages which were partly fortified. BaKwele religion was expressed primarily in witchcraft beliefs and in several more or less communal rites of intensification which, incidentally, decided or validated the social status of the patriline. Numerous masks were used in a major ritual complex performed on irregular occasions, on the average perhaps every other year; a few masks were occasionally used in other ways. European contact with the BaKwele began only at the very end of the last century and eventually, probably during the 1920s, led to abandonment of the rituals in which the masks were used. Some carvers have continued to make masks, however, at the request of administrators and for local use on public occasions—for entertainment rather than ritual. The younger generation have French as a second language and are commonly literate in it. Acculturation has been rapid and apparently welcomed, at least during the last few decades, but there is little sign of interest in European art. There seems no reason to believe BaKwele evaluations of their own masks would have been influenced by contact with European traditions.

The field worker took with him photographs of all the masks known to him, from publications and from museum and private collections, which might be presumed to have been made by the BaKwele or by neighboring peoples. The photographs were 39 in number and were of uniform size—5" × 7"—but unfortunately not of uniform quality; some were made from published illustrations of good quality or photographs made under good conditions, but some were from poor published illustrations or from snapshots made under unfavorable conditions, and one was a photograph of a drawing.

The 39 masks, as represented in the photographs, were judged for their esthetic value—i.e., for how good they were as works of art—by thirteen experts (advanced art students and others able to make such judgments) in New Haven, Connecticut. As nearly as BaKwele language and general culture permitted, the field worker tried to obtain judgments by similar criteria from individual BaKwele; there was a difference, as will be seen.

FIELD PROCEDURE

The photographs were used as a point of entry into the ethnological study of masks. For this reason as well as for the present purposes, elders

were sought as subjects. The making of masks had been confined to men, and their traditional use had been practiced only by men now well over 50; consequently women were not interviewed, and emphasis was placed on interviewing older men.

At strategic times the 39 photographs were set out in random order on a large table or a mat on the ground. The field worker used an interpreter with whom he spoke in French. The interpreter was requested to ask the subject to choose the masks which he found "*les plus beaux*" or "*qu'il aime plus que les autres*." The terms used by the interpreter in carrying out this instruction referred to goodness in general or to beauty as it applies to persons. The BaKwele constructions for "well-made" or "well-carved" were not used in posing questions but were sometimes used by the interviewees in their comments on the masks.

The subject was asked to choose from the 39 the four masks he considered to be the best. After they had been chosen, he was asked to choose four more, and then to continue choosing in this manner until the last photograph had been judged. Four of the subjects were not willing to continue all the way through, and the photographs they left at the end were all given an identical score based on the average of all the categories not used. One of these four had also selected in groups of eight rather than four, and a similar procedure of giving tied scores was followed.

The conditions under which data were gathered were in several respects unfavorable for the purposes of the study. For one thing, the circumstances were often hectic. The subject was often surrounded by all members, human and animal, of his hamlet. Children would seize photos from the table and attempt to carry them off. Women would pick up the photos they preferred and hand them to the subject for his approval. Such attempts to influence the subject were inevitable and sometimes intensive, but the field worker's position as guest of the settlement precluded any remonstrances which might have been effective. Attempts to conduct the test in seclusion seemed to make for awkwardness, as it forced upon the men the disagreeable task of having to keep their women and children out of the house in which the test was given; such attempts were therefore abandoned. In any event, the sixteen different subjects were from different hamlets, and in only two instances was one of them present when another was making his judgments; in each of these the former had made his judgments previously.

The physical arrangements themselves were less than satisfactory. Presenting so many photographs at one time made for dubious care in inspecting the entire field and also made the prints more likely to be blown by the wind or to appear in widely varying illumination.

Because photographs rather than actual masks were used, the judgments cannot be confidently asserted to be equivalent to those which would have been made of the masks themselves. To be sure, this is true of the New Haven experts as well. But the New Haveners are at least used to seeing both an object and a photographic representation of it, so their imagination may generally carry them fairly well from the latter to the former. This is much less true for the BaKwele subjects, particularly since they were mostly

older men who had been less exposed to the printed page and accompanying illustrations than had younger members of the group. In particular, some photographs in which strong shadows were apparent were subject to misinterpretation, for the shadows might be thought to be a part of the mask by a person not used to interpreting photographs as representations of three-dimensional objects.

RESULTS

The ratings made of each photograph by each of the sixteen BaKwele are presented in Table 1, each judge being identified by a number, age, and one of three categories of relation to masks. First come four judges who were carvers. (Two of these, Nos. 3 and 4, were old-time mask carvers. No. 2 was a carpenter who carved masks in the traditional style for the field worker. No. 1 has produced ivory carvings—not masks—for sale to administrators, missionaries, and business men.) Then come four judges who as cult leaders had used masks in the traditional manner. Finally appear eight other BaKwele interested in or knowledgeable about masks. An entry of 9 means that the judge placed a photograph in the topmost of ten possible groups of four as being one of the four best in his estimation. An entry of 0 means that he placed a photograph among the four poorest. In this table the 39 photographs have been arranged in descending order of their evaluation by the New Haven experts. It is immediately apparent that there is some tendency for the BaKwele judges to agree with the New Haven judges, as high ratings are more frequent in the upper part of the table. But we need to measure the degree to which agreement is present and the confidence that may be placed in the evidence of agreement. For these purposes certain standard statistical procedures have been followed.

First, to measure the evaluation of the photographs by the consensus of any particular group of subjects, factor analysis has been employed. For example, correlation coefficients were obtained to express the degree of similarity between the ratings of each possible pair out of the thirteen New Haven judges. A factor analysis was performed to determine the extent to which each of the New Haven judges expressed a tendency shared by their group as a whole, retaining only the single first factor which would represent this consensus, and then factor scores were calculated for each of the 39 photographs to express their standing in relation to this factor. These factor scores represent the average rating given to a photograph by the thirteen New Haven judges when each one's ratings are weighted in proportion to the extent to which his judgments correlate with the general consensus of the thirteen. Identical procedures were followed for the four BaKwele who were themselves carvers, for the four BaKwele who were cult leaders, and for the eight other BaKwele considered together as a single group. The evaluations that result for each photograph are also presented in Table 1.

The factor scores representing evaluations by the thirteen New Haven experts are the point of reference for the calculations which then followed. To determine how closely the consensus of the BaKwele carvers agreed with

TABLE 1
39 Mask Photographs: Identification, Ratings, and Factor Scores

Id. No.	Source ^a	Ambig.	Foreign	Play	Fierce	Ratings by BaKwile																Factor scores representing evaluation by consensus of				
						Carvers				Cult leaders				Others								Carvers	Cult leaders	Other BaKwile	All 16 BaKwile	N. H. experts
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	1.0	1.7	4	9	1.7						
11	Plaza, 1956, no. 24-A					7	8	4	9	9	9	3	7	7	6	8	9	3	5	6	2.5	1.0	1.7	1.7	1.7	
13	Fagg, 1953, no. 104, pl. xxxiv					8	4	6	7	1	1	3	5	2	3	4	7	3	4	2	2.5	1.2	1.7	1.6	1.6	
29	Plaza, 1956, no. 24-D					9	7	9	8	3	8	3	5	7	3	3	6	3	2	4	2.5	1.3	1.5	1.4	1.4	
32	Kamer & Kamer, 1957, no. 193					9	9	8	8	4	4	3	7	4	2	2	4	3	9	9	6.5	1.1	1.0	2.0	1.3	
14	Sagy, 1952, fig. 184					3	6	6	1	4	4	3	7	4	2	2	4	3	2	4	2.5	1.7	2.1	2.0	1.3	
38	Cloutot & Level, 1925-26, pl. vii					7	9	9	7	9	9	9	9	9	8	8	8	8	9	8	2.5	1.4	2.0	1.8	1.2	
13	Mus. Prim. Art (N. Y.), 1961, no. 70					7	7	9	9	9	9	9	9	9	9	9	9	9	9	8	2.5	1.5	2.3	2.0	1.2	
20						5	5	2	2	8	5	3	5	9	4	2	1	3	5	0	8	2.5	1.6	1.8	1.2	
35	Pepper, 1958, p. 36					5	5	2	2	8	5	3	5	9	4	2	1	3	5	0	8	2.5	1.3	2.0	1.1	
28	Eth. Mus. (Gothenburg), p. 5, upper R					5	5	2	2	8	5	3	5	9	4	2	1	3	5	0	8	2.5	1.3	2.0	1.1	
2	Sweeney, 1935, no. 414					5	5	2	2	8	5	3	5	9	4	2	1	3	5	0	8	2.5	1.3	2.0	1.1	
34	Olbrechts & Claerhout, 1956, no. 183					5	5	2	2	8	5	3	5	9	4	2	1	3	5	0	8	2.5	1.3	2.0	1.1	
33	Mus. Prim. Art (N. Y.), 1961, no. 71A					1	1	7	1	0	2	4	3	5	2	2	4	3	5	0	8	2.5	1.3	2.0	1.1	
25	Eth. Mus. (Gothenburg), p. 7					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.3	2.0	1.1	
31						4	2	8	9	9	9	9	9	9	9	9	9	9	9	8	2.5	1.4	1.7	1.5	1.5	
27	Eth. Mus. (Gothenburg), p. 5, upper L					9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	2.5	1.4	1.7	1.5	1.5	
7	Eth. Mus. (Gothenburg), p. 8, L					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
24						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
26						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
8						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
5	Kjelberg, 1957, p. 164					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
19	Mus. Prim. Art (N. Y.), 1961, no. 71					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
36						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
18	Ass. Pop. AmiaMus., 1948, fig. 32, p. 59					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
17	Cloutot & Level, 1925-26, pl. xxi					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
4						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
1						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
21						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
15	Eth. Mus. (Gothenburg), p. 5, lower L					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
10	Eth. Mus. (Gothenburg), p. 4					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
9	Conard, 1934, p. 713					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
22	Eth. Mus. (Gothenburg), p. 5, lower R					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
6	Eth. Mus. (Gothenburg), p. 8, R					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
23						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
39	Burrsena, 1960, fig. 7					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
37	O'Reilly's (N. Y.), 1953, no. 129					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
16	Cottes, 1911, pl. xxxiii					7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
12						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5
30						7	7	2	2	5	1	3	5	9	4	2	1	3	5	0	8	2.5	1.4	1.7	1.5	1.5

the New Haven consensus, for example, the factor scores for the consensus of the four carvers were correlated with the factor scores for the New Haven experts. The same procedure was followed for each of the other groupings of BaKwele judges. Also, the evaluations by each individual BaKwele judge were similarly correlated with the consensus of New Haven evaluations. Results for all 39 photographs together are presented in the first columns of Table 2. For each of the groupings of BaKwele judges, the evidence of tendency to agree with the New Haven judges is significant at the 1 per cent level (that is, so great an appearance of agreement, in the absence of true agreement, could arise by chance less than 1 per cent of the time). The evidence of tendency to agree with the New Haven judges is significant at least at the 5 per cent level for each individual carver and each individual cult leader, and also for four of the eight other BaKwele.

TABLE 2

BaKwele Evaluations of Mask Photographs: Resemblance to
Evaluations by Art Experts in New Haven, Connecticut
(Each entry is a correlation coefficient expressing the degree to which a
given set of BaKwele evaluations resembles the New Haven evaluations.)

BaKwele Judges	All 39 Photographs	Photographs remaining after 8 poor ones removed			
		All 31 Remaining Photographs	All 27 Remaining Photographs	16 Play Masks	8 Fierce Masks
Consensus of					
a. 4 Carvers	.48**	.50**	.43*	.48*	.85**
b. 4 Cult leaders	.44**	.51**	.51**	.55*	.54
c. 8 Others	.38**	.42**	.34*	.45*	.37
d. All 16 judges	.44**	.48**	.42*	.50*	.76*
Individual judges with ages					
Carvers					
1. 35-38	.49**	.46**	.34*	.31	.80**
2. 40	.46**	.50**	.44*	.48*	.52
3. 65	.33*	.34*	.33*	.44*	-.05
4. 65	.30*	.39*	.36*	.29	.62
Cult leaders					
5. 55-60	.39**	.37*	.35*	.23	.64*
6. 55	.36*	.43**	.53**	.60*	—
7. 65	.38**	.45**	.40*	.34	.52
8. 55-60	.28*	.39*	.37*	.57*	-.08
Others					
9. 17-18	.40**	.39*	.31	.34	.25
10. 42	.32*	.33*	.21	.28	.68*
11. 23	.47**	.52**	.46**	.68**	.09
12. 65	.15	.18	.19	.15	—
13. 50	.22	.31*	.23	.45*	.30
14. 65-70	.14	.13	.06	.25	-.64
15. 45-50	.36*	.35*	.42*	.48*	.51
16. 50	.25	.36*	.26	.14	.48

* = probability of so large a positive correlation with true correlation zero, <.05

** = probability of so large a positive correlation with true correlation zero, <.01

When eight photographs which posed ambiguities for inexperienced viewers (mostly because of shadows which might be confused with the masks) are removed, the results for the 31 remaining are more striking still, as may be seen in the second column of Table 2. The correlations with the New Haven judgments are somewhat larger, and there are now only two of the sixteen BaKwele judges for whom the evidence of agreement with New Haven judges fails to be significant at least at the 5 per cent level.

Of these 31 remaining photographs, four are among those stated by BaKwele judges to be foreign (in addition to those foreign masks which were among the eight offering photographic difficulty). The evidence of agreement with the New Haven judges, for the 27 photographs that remain after removal of those four, is less decisive although still strong, as shown in the third column of Table 2. Among these 27 BaKwele masks there are sixteen "play" or "entertainer" masks. These were used in dancing which heightened and sustained ceremonial occasions, and to them the BaKwele would apply words which might be translated by "beautiful" or "good" but which would denote power, protection, and opposition to evil. Another eight of the masks were of a type considered to be fierce and fearsome, used occasionally to terrorize the village. Results calculated separately for masks of each of these two types are presented in the last two columns of Table 2. Although the results are again less decisive than for the entire 39 photographs, there is still very convincing evidence of a tendency toward agreement with the New Haven judges despite the small number of masks that now remain.

In the statements of statistical significance mentioned thus far and noted in Table 2, the particular group of judges is taken as a given, and the question is what generalization can be made about how these judges would theoretically have responded to an infinitely large population of masks. But suppose we instead take a set of masks as given and ask with what confidence one can generalize to the population of people of whom our judges are representative. In this event a very decisive answer is given by the nearly uniform positive direction of the correlations presented in Table 2. Only when the number of masks is reduced to eight—the fierce masks dealt with in the last column of the table—is there ever a negative correlation. In every other instance the direction of correlation indicates some tendency for agreement rather than disagreement with the New Haven consensus. The probability that this unanimity of direction could have arisen by chance is, for the four carvers or four cult leaders separately, $1/16$; for the carvers and cult leaders together, $1/256$. Finally, for all sixteen judges together the probability of their independently yielding positive rather than negative correlations by chance is something less than 1 in 32,000.

We have thus far stressed the overwhelming evidence of a tendency toward trans-cultural agreement because we think this is the aspect of the results that will be surprising to many people. Careful inspection of the data in Table 1, however, will indicate that an equally striking result is confirmation of the fact that the BaKwele also have evaluative standards not shared with New Haven experts. Photographs 14 and 20, for example, stand out as instances of remarkably close agreement among the sixteen BaKwele

judges in placing the photographs at or near the top of the list, a position higher than they reach in the consensus of the New Haven judges. The photographs that are second and third in the New Haven consensus, on the other hand, are not given the maximum rating of 9 by a single one of the sixteen BaKwele judges. Very clearly there are determiners of agreement among the BaKwele that have nothing to do with the criteria underlying the consensus of the New Haven experts.

This fact may be represented in summary by asking how well the consensus of one subgroup of the BaKwele judges agrees with the consensus of another in comparison with the extent to which it agrees with the consensus of New Haven judges. If there were no special tendency for the BaKwele to agree with each other, one would expect the correlations among subgroups to be lower than the correlations between each subgroup and the New Haveners, since the consensus of the New Haveners is based on a larger number of subjects and is therefore a more stable ordering. The facts are quite the other way. For all 39 masks considered together the consensus of the four carvers has a correlation of .84 with the consensus of cult leaders and of .93 with the consensus of eight other BaKwele, while the latter two have a correlation of .86 with each other; all these are much higher than the correlation between any one of these consensi and that of the New Haven experts, presented in Table 2.

The same kind of result is obtained with each of the smaller groupings of masks reported in Table 2 with the sole exception of the fierce masks; in that instance the consensus of the carvers has a correlation of .59 with that of the cult leaders and .52 with that of the other BaKwele, and the latter two groups have a correlation with each other of only .10, whereas the correlations of subgroups with the New Haven consensus reach as high as .85. Since there were only eight masks of this kind, it is impossible to be sure whether we have here an instance where there is no special cultural consensus, or whether sampling error is responsible. The customary use of this type of mask is such, however, as to favor the supposition that there may be a genuine lack of agreement because of the intrusion of other sources of evaluation. An attitude of fear and therefore dislike might well characterize those people whose experience had been that of being inconvenienced or harmed by the use of such masks, whereas those who controlled the masks or were their primary users might well have a positive feeling because of the power associated for them with such masks.

DISCUSSION

Why do the New Haven experts and the BaKwele judges show some agreement in their evaluations of BaKwele masks? Perhaps because the masks do really vary one from another in general suitability for arousing and sustaining interest in anyone who enjoys visual art, and both sets of judges are sensitive to this variation. We find this interpretation plausible in the light of our own view of the masks: the ones rated high by both New Haveners and BaKwele do indeed seem unusually beautiful, and more in-

teresting with repeated looking; we feel the same way about some of the masks with lower ratings, but not about all of them. To say what it is in the masks that leads us to react this way would not be easy. The history of esthetic theory suggests that the relevant characteristics of the masks are many and complex.

But is such an interpretation necessarily called for? Perhaps the agreement between New Haveners and BaKwele has a much more superficial origin. Perhaps it is based entirely on some fairly simple and obvious characteristics of the masks. We have looked at the photographs with this question in mind and can see simple characteristics that may play some part. Two of the masks (Nos. 12 and 16) rated very low by both groups appear in very poor reproductions, enlargements from half-tone illustrations with conspicuous grain; this inadequacy of reproduction is shared by only one of all the other photographs (No. 17). This poor quality as photograph may well have rendered more extreme an evaluation which would in any event have been low for Nos. 12 and 16 in view of the nature of the masks themselves. Another characteristic which may have influenced both sets of judges is the quality of workmanship; careful smoothing is apparent more frequently, for example, in masks rated high. Here, too, this is not likely to have been the sole basis of judgment. Carefully finished pieces that rank high seem to have other more complex esthetic attributes as well. And there are exceptions; for example, No. 37, ranked very low by both groups, seems to have been carefully finished (although subsequently damaged). Our inclination is to feel that simple features such as these are not the sole basis for agreement, but the information provided by this one study is not sufficient to settle the question.

Another possibility, of course, is that European influence is after all a probable explanation for the agreement, despite our judgment to the contrary. This explanation might be supported by the fact that the few younger BaKwele judges, as may be seen in Table 2, on the average have closer agreement with the New Haven experts than do the older BaKwele judges. Such a relation to age; if it may be inferred from so few cases, may, however, have quite a different meaning; younger men who seemed interested enough in masks to be chosen as judges may be more likely to have an esthetic interest, since so many of the earlier reasons for interest in masks have disappeared. Here, too, the information provided by this one study cannot be decisive.

Inspection of the masks may also provide clues about how the standards of the two sets of judges differ from one another. Of the four masks which show the greatest tendency to be rated higher by New Haven experts than by BaKwele, two are perhaps in poor condition (No. 29 with kaolin adhering only in patches and No. 26 with tusks missing where the BaKwele would expect them), and two are of the fierce type (Nos. 26 and 38). One (No. 3) is an ambiguous photograph both because of shadows and because the mask is in a slanting position. These characteristics of mask or photograph are all somewhat more frequent here than among the rest. A reasonable explanation is that the New Haven judges are more used to admiring

damaged works of art, less repelled by (partly because not knowing about) fearsome qualities or associations, and better able to interpret correctly ambiguous photographs. On the other hand, all four masks which show the greatest tendency to be rated higher by the BaKwele than by the New Haveners (Nos. 1, 10, 27, and 30) are somewhat deviant from, while still within, BaKwele tradition. This is true of only a few of the other masks, and it may be that BaKwele are attracted by such deviance, whereas the New Haveners could probably in no case have been aware of it. These four masks also share the characteristic of having some very salient feature—either horns, wings or pronounced light-dark contrasts. This is true of more than half of the other masks, too, but it seems possible that the attention of the BaKwele judges, in view of their lack of experience in dealing with photographs, was especially readily caught by such salient features.

Consistencies of preference, more or less distinctive to a society, are obviously to be expected. The method used here could be of value to field workers who wish to identify clearly the preferences characteristic of the community they are studying, in order to relate those preferences to other aspects of the culture. Preference judgments on the same materials need to be obtained in two or more different groups, however, before one can even begin to have a satisfactory basis for judging in what respects the preferences are distinctive to one group and in what respects they are shared by different groups.

SUMMARY

Photographs of BaKwele masks were judged for esthetic merit by art experts in New Haven, Connecticut. During a field trip to the Congo Republic (Brazzaville), judgments of these same photographs were obtained from sixteen BaKwele men interested in or knowledgeable about their masks. The consensus of the sixteen BaKwele and of subgroups of them, and most of the sixteen individuals, showed significant agreement with the consensus of New Haven experts. The finding of some trans-cultural agreement cannot be interpreted confidently from this one study alone; it is consistent, however, with the notion that the esthetic appeal of a work of art to an art-involved viewer is partly a function of universals of human nature, and it should encourage further trans-cultural comparison of evaluative responses to art. The BaKwele also showed agreement among themselves on other bases than those shared with the New Haveners.

NOTES

1. The field study on which this article is based was made under a Ford Foundation Foreign Area Training Fellowship. However, the conclusions, opinions, and other statements in this article are those of the authors and not those of the Ford Foundation. The analysis of the data was done under Cooperative Contract No. 1748, U. S. Office of Education.
2. Masks for which no source is given are in public or private collections and have not, to our knowledge, been pictured in publications.

BIBLIOGRAPHY

- Association Populaire des Amis des Musées (Paris). 1948. Le Musée Vivant: Numéro Spécial.
- Burssens, H. 1960. Enkele Zanda-Masker uit Uele. Congo Tervuren 6: no. 4.
- Child, I. L. 1962. Personal Preferences as an Expression of Aesthetic Sensitivity. *Journal of Personality* 30: 496-512.
- Clouzot, H., and A. Level. 1925-26. Sculptures africaines et océaniques. Paris.
- Cottes, A. 1911. La Mission Cottes au Sud-Cameroun. Paris.
- Cunard, N. 1934. *Negro Anthology 1931-1933*. London.
- Ethnographic Museum (Gothenburg). 1955-56. Årstryck.
- Fagg, W. B. 1953. The Webster Plass Collection of African Art. London.
- Firth, R. 1951. *Elements of Social Organization*. New York.
- Kamer, H., and H. Kamer. 1957. *Arts d'Afrique et d'Océanie*. Cannes.
- Kjellberg, S. T. 1957. Masker ur Benkt-Åke Benktssons Samling. *Kulturens Årsbok*: 157-184.
- Kluckhohn, C. 1955. Ethical Relativity: *Sic et Non*. *Journal of Philosophy* 52: 663-677.
- Lawlor, M. 1955. Cultural Influences on Preferences for Designs. *Journal of Abnormal and Social Psychology* 61: 690-692.
- Linton, R. 1954. The Problem of Universal Values. *Method and Perspective in Anthropology*, ed. R. F. Spencer, pp. 145-168. Minneapolis.
- Lowie, R. 1921. A Note on Aesthetics. *American Anthropologist* 33: 170-174.
- McElroy, W. A. 1952. Aesthetic Appreciation in Aborigines of Arnhem Land: A Comparative Experimental Study. *Oceania* 23: 81-94.
- Museum of Primitive Art (New York). 1961. *Traditional Art of the African Nations*.
- Olbrechts, F. M., and A. G. Claerhout. 1956. *Het Masker*. Antwerp. (Catalogue of an exhibition at the Royal Museum of Fine Arts, Antwerp, Sept. 16-Nov. 15, 1956).
- O'Reilly's Plaza Art Galleries (New York). 1953. Catalogue 3432, sec. 3.
- Pepper, H. 1958. *Anthologie de la vie africaine* (booklet accompanying record album of same name). Paris.
- Plass, M. 1956. *African Tribal Sculpture*. Philadelphia.
- Segy, L. 1952. *African Sculpture Speaks*. New York.
- Sweeney, J. J. 1935. *African Negro Art*. New York.

SOME TRANSCULTURAL COMPARISONS
OF ESTHETIC JUDGMENT*¹

American University of Beirut and Yale University

C. S. FORD, E. TERRY PROTHRO, AND IRVIN L. CHILD

A. INTRODUCTION

If people in Western culture who have some expert acquaintance with art are asked to compare the esthetic values of works of art, substantial agreement among the judgments is likely to be found [see Child (1)]. Does the agreement result from the fact that such people share certain conventions of current Western culture? Or does it arise partly because each person considers the works of art carefully enough to find in his interaction with them much the same values (positive and negative) that an art expert in some different culture looking at the same works would also find? Is esthetic evaluation culture bound? Or does it have some degree of universality because it is partly based on universal tendencies in human nature?

The kind of evidence needed to settle this question is, of course, sets of judgments on the same works of art made by experts of radically differing cultural background. No evidence of this sort seems to have been available in the past, but steps toward assembling such evidence are made in this paper and in one by Child and Siroto (3).

B. PROCEDURE

Child, in connection with other research (2) obtained esthetic evaluations by experts in New Haven, Connecticut, of various sets of reproductions of works of art; for example, 60 postcard-size reproductions of abstract paintings were judged for relative artistic merit. Each judge divided the paintings into 10 piles of about six cards each—from the six best to the 6 poorest. A set of 32 reproductions of Bambara antelope headpieces were similarly evaluated by being divided into 10 piles arranged according to a judgment of relative esthetic merit. With smaller sets, such as 12 still-life drawings by an artist who attempted to produce for the project drawings of similar subject matter

* Received in the Editorial Office, Provincetown, Massachusetts, on July 7, 1965, and given special consideration in accordance with our policy for cross-cultural research. Copyright, 1966, by The Journal Press.

¹ The analysis of data and preparation of this report were supported by Cooperative Contract Number 2840 with the U. S. Office of Education.

and style but that differed radically in esthetic value, judges were asked to rank order the items. Rating or ranking procedures were then applied to a number of such sets of pictures altogether. In view of the plans of Ford and Prothro to engage in field work in regions where there might be art experts who had had little or no contact with evaluative customs in the main stream of Western art, some of the materials just mentioned were utilized to obtain judgments in the field. From each of several sets that offered promise that the pictures would be fairly well understood regardless of the cultural background of a viewer, a trio of pictures was chosen on which the New Haven judgments seemed to justify the hope that if the three objects were compared directly one with another, New Haven experts would agree on the order of merit. With two exceptions, however, it did not prove possible to use identical trios on both field trips.

The materials, as they were judged in the field, differed in two ways from the materials originally judged in New Haven. First, in some instances the reproduction was substantially different in ways that might influence esthetic judgment (e.g., original drawings *vs.* photographs). Second, the three pictures in a set were judged as a group of three; whereas, earlier they had been embedded in a larger set of similar objects to be judged. Therefore, after the field work was complete, judgments were made in New Haven in a manner comparable with those made in the field. On this second judging in New Haven, the trios alone were presented as they had been in the field; moreover, the judging was done with the identical reproductions used in the field or with reproductions not discriminably different. In the second New Haven judging, the authors required a higher standard of expertness of the people to make the judgments. For the initial judging of the larger sets, the judges mostly had been students in the Yale Art School or graduate students in the History of Art. In the second New Haven judging, the judges mostly were professional artists or teachers of art; students were included only if they were known from earlier performance as esthetic judges to show a high tendency toward agreement with the consensus of expert judgment.

When materials identical with those used in the field were thus judged in trios by New Haven experts, the order of merit in which the three were placed usually remained the same as it had been when the three were part of a larger set evaluated by the original judges. There were, however, some exceptions in which the work originally in the middle of the three was now at one extreme or the other. Unanimity was usually not found either in the original or in the later judging, even about the relative ranking of the best and poorest pictures in relation to each other; and in some instances the dis-

agreements were fairly numerous. In view of the varying degree of agreement among the New Haven judges, we decided to report the results in two different ways: one way is to deal only with the instances of good agreement among New Haven judges; the second way, to include all the data collected.

1. First, to see whether or not good agreement among New Haven judges is confirmed by the judgment of groups in other cultures, we restricted the materials to those trios in which the works ranking first and third according to the original New Haven judges remained in those positions when the New Haven judges rated the trio alone. We have considered for these trios only the pair consisting of the highest-ranking and lowest-ranking work and have used these only if at least 12 out of the 15 new New Haven judges agreed with the original New Haven judges in the order in which they placed these two in relation to each other.

2. In order to see the results when all data collected are utilized, we have considered the evaluation of each picture by all the judges in a foreign group and have compared this evaluation with the evaluation by all the new New Haven judges who saw the same trios.

1. *Data from Fiji*

One of us (Ford) took 11 trios of pictures with him on a field trip to Fiji. There, he showed them to five men and one woman on Naviti, one of the westerly group of islands known as the Yasawas. The people of this island have been relatively isolated from influence by Westerners or by the Indians, who have become an important part of the population in much of Fiji. In recent years, some have made frequent visits to Lautoka on Viti Levu, but these trips are brief and primarily for the purpose of taking produce to the big island and bringing back such items as tobacco and sugar. Although members of the younger generation are taught English in school, they do not use English as a medium of communication. Most of the people on Naviti can read and write in Fijian, but their reading matter consists primarily of letters from friends and relatives and the Bible.

A major source of possible Western influence is the radio. Each village on Naviti has one or two radios which are constantly being played, but the people listen mostly to music and to the broadcasts in Fijian. Another source of Western influence results from the fact that several young Fijian men from the Yasawas served with American armed forces in the Solomons during World War II. The ones with whom this experience was discussed, however (although apparently having liked the Americans they knew), much preferred

the Fijian way of life and had rejected much of what they had learned of the American way of life.

In general, then, the nature and the amount of Western contact do not suggest that the esthetic standards of art specialists in the Western world are likely to have been communicated to or adopted by the Fijians resident on Naviti.

The men interviewed were competent house builders and knew how to carve wooden kava bowls and how to make coconut-shell cups. No other wood carving is done on Naviti now. People still have some of the war clubs and dance paddles carved in former times and handle them frequently during dances. The woman interviewed was a skilled mat and basket maker and tapa designer. In these respects the persons interviewed are typical of the community, and they are also typical in being to some degree artistic creators. Within the traditional style of making the objects appropriate to one's sex, there is decided freedom for individual variation. The merits of objects locally produced are discussed sometimes with reference to skill, but evaluations of art in other terms do not seem to appear in conversation. It was easy, however, to get Fijians to express preferences because the rating or the ordering of things and people by various criteria is a congenial mode of thought among them. But it was difficult to get an individual alone, so that his own private judgment could be obtained without being influenced by or influencing the judgments of others. The persons contacted would have preferred to discuss and express their opinions in a group. It is largely because of this difficulty that so few judges were obtained since we wished each set of judgments to be independent.

The instructions were to indicate the order of preference among the three pictures presented at a time. Characteristically, the judge studied the three intently for a bit and then, with no hesitation or wavering, indicated his order of preference by pointing in immediate succession to his first, second, and third choices. In conversation afterward, it appeared that one of the things influencing the judgment of the men—especially with regard to the masks—was the work and skill they judged to be required to make the masks.

Of the 11 trios of material shown to the Fijian judges, there are six that meet the criterion stated earlier for good agreement among New Haven judges on the best and the poorest of the trio. If we consider the six Fijian judges as a jury expressing Fijian opinion on the resulting pairs of works of art, we obtain the following results: for one pair, three out of the six Fijians agreed with the consensus of the New Haven judges; for one pair, four out of the six agreed with the New Haven consensus; for each of three pairs, five out

of the six agreed with the consensus of the New Haven judges; and for the sixth pair, the six Fijian judges were unanimous in their agreement with the New Haven consensus. If we treat these proportions as deviations from the chance value of .50, the average value of .78 differs from .50 with a t of 3.6, significant beyond the .01 level (with a one-tailed test).

If we regard the six pairs as a test applied to individual Fijian judges to measure how well they tend to agree with New Haven judges, we find that the six Fijians happen to distribute themselves in exactly the same way as the six pairs, so that we may generalize to a population of judges with exactly the same confidence as to a population of pairs.

If we consider all the pictures used, agreement between representatives of the two cultures is still impressive. Within each trio, each Fijian judge ranked each work as first, second, or third; and these rankings were added over the six judges. The rankings obtained from the New Haven judges were summed in the same way. The correlation between these two summed rankings over the 33 works of art is .56. As the values are interdependent in sets of three, there appear to be 20 degrees of freedom; and for this number of degrees of freedom, a correlation of .56 is significant at the .005 level (by a one-tailed test).

From the instances of excellent agreement with the New Haven judges, and from all the data considered together, there is significant evidence that the Fijian judges tend to agree with the New Haven judges in their evaluative responses to works of art.

2. *Data from Greek Islands*

Another of us (Prothro) took five trios of pictures to the Western Cycladic Islands in Greece and showed them to four local craftsmen who did creative work but who were little acquainted with the outside world of art. Two were women (one, a weaver; the other, a painter of ceramics). The other two were men (one, a pottery maker; the other, a basket maker who made objects such as purses and fans).

The islands of Serifos and Sifnos, where these craftsmen live and work, are far from the principal tourist routes. They are a few miles apart and are served by two weekly steamers out of Piraeus, as well as by fishing boats and small caiques.

Some natives of these islands have emigrated to Piraeus, and some Athenians have traditionally had their summer homes there; so the islands have ties with the mainland. In each place enquiries were made about native craftsmen, and an interview was sought with those who seemed to be of highest repute. The

field worker spent more than a week on the islands on each of four occasions and came to know the craftsmen. One of the craftsmen insisted on being joined in a long session of "testing the new wine" before he would (on a subsequent occasion!) judge the trios. Each of the four craftsmen interviewed had learned his skill as a child and had practiced it since adolescence. All were between 30 and 40 years of age. They seemed to have no knowledge of contemporary art in the outside world; for example, they had not heard of Picasso. They did their work in a traditional manner, but had won their reputations by skill and pleasing original variations. None seemed to be acquainted with theories or books about art. Conversation indicated that they had no abstract concept of "art" as a field, but thought of it as goodness in craftsmanship (Greek: *kalotechnika*).

All four of the craftsmen were literate, with a few years of primary education. They seemed to have ready access to Athenian newspapers, and the two women were interested in poetry and had some books of poetry in their homes. All were quite poor, however; and no magazines or illustrated books were in evidence in their homes. The only works of art visible, other than objects of practical use, were icons which had only modest claim to beauty. As is often true of peasants in the modern age, the persons interviewed loved to talk and discuss daily events, but the events largely were of a local nature or were political topics heard on the radio. Except for military service for one of the men, no one had traveled farther than Athens, but the men had been to Athens perhaps as many as 10 or 12 times. The women had been perhaps three or four times. They stayed with island friends during their few days in Athens and had not seen any art museums. Their products are sold through middlemen from the city who are friends or relatives of the islanders. Sales are on a basis of orders for duplicates of a sample, and the craftsmen express their originality in designing the samples. Each had probably been influenced by seeing the artifacts brought by summer visitors and by discussing orders with those visitors.

Of the five trios shown, only three met the stringent criteria described for good New Haven agreement on the best and the poorest of each trio. For one of these pairs, two of the four judges agreed with the New Haven consensus; for another, three agreed; for the third, all four agreed with it. The proportion of agreement with New Haven judges is almost as large as for the Fijians (75 per cent, in comparison with 78 per cent), but with so small a number of pairs the deviation from the chance value of 50 per cent is not statistically significant ($t = 1.73$, $df = 2$). The numbers of pairs for which the Greeks agreed with the New Haven consensus were (for the four

artists) one, two, three, and three (out of the three available). The results, however, are not statistically significant ($t = 1.57$, $df = 3$).

When all of the 15 pictures judged by the four Greek artists are considered, the correlation between the summed rankings and those given by the new New Haven judges to the same pictures is positive, but it is too small to be statistically significant ($r = .28$, $df = 8$).

The small body of evidence from these four Cycladic Islands artists, then, suggests a tendency for esthetic judgments to agree with those made by experts in New Haven; but the evidence is not statistically significant.

On the same field trip there was opportunity to show the same pictures to two artists who do have some sophistication in the art world: a woman artist of Thessaly and a man who is an interior decorator there. Each of these two agreed with the New Haven experts on every one of the three pairs on which the New Haven judges were well agreed. When the evaluations by these two artists of all 15 pictures used are compared with those obtained from the New Haveners, the correlation is .86.

C. SUMMARY AND CONCLUSION

Six residents of a remote Fijian village (all engaged to some extent in craft-work) and four craftsmen in the Cycladic Islands of Greece expressed preferences within trios of works of art, which were also evaluated for relative esthetic merit by Americans greatly interested in art. Significant evidence of agreement was found between Fijians and Americans. Greek craftsmen also tended toward agreement with the American judges, but the tendency was not statistically significant. Taken together, these two sets of data strongly suggest a tendency toward transcultural agreement in the evaluation of works of art, by people with an active interest in art. Previous studies (4, 5) which have looked for transcultural agreement in esthetic preference between people not selected for interest in art and failed to find it must be re-evaluated in the light of the present data. Esthetic responses to works of art may be made only by some people in most communities (perhaps by everyone if favored by the culture), yet show some consistency wherever made.

REFERENCES

1. CHILD, I. L. Personal preferences as an expression of esthetic sensitivity. *J. Personal.*, 1962, 30, 496-512.
2. ———. Personality correlates of esthetic judgment in college student. *J. Personal.*, 1965, 33, 476-511.
3. CHILD, I. L., & SIOTO, L. BaKwele and American esthetic evaluations compared. *Ethnology*, 1965, 4, 349-360.

4. LAWLOR, M. Cultural influences on preferences for designs. *J. Abn. & Soc. Psychol.*, 1955, **61**, 690-692.
5. McELROY, W. A. Aesthetic appreciation in aborigines of Arnhem Land: A comparative experimental study. *Oceania*, 1952, **23**, 81-94.

Department of Psychology
Yale University
333 Cedar Street
New Haven, Connecticut

COMPARISON OF ESTHETIC JUDGMENTS BY AMERICAN
EXPERTS AND BY JAPANESE POTTERS*¹

Keio University and Yale University

SUMIKO IWAO AND IRVIN L. CHILD

A. INTRODUCTION

Does a person who has developed some expertness in art within a particular cultural tradition have evaluative standards that resemble those of experts in other cultural traditions? or are esthetic evaluations entirely relative, varying arbitrarily from one culture to another? Many psychologists, social scientists, and art experts (influenced by the general position of the cultural relativism so prevalent in recent times) take it for granted that esthetic evaluations are completely culture bound. On the other hand, most theories of esthetic value (while not explicitly facing the problem of cultural variation) imply the existence of universally valid esthetic standards.

The results of two small-scale studies (1, 2), which encourage the belief that there is some element of universality in esthetic evaluation, had become available when one of us (Dr. Iwao) was planning to return to Japan after several years of study in the United States. By this time the other (Dr. Child) had prepared a larger quantity of suitable stimulus materials than had been available at the time of the earlier studies. The authors decided to take advantage of the opportunity to see whether or not people working in the arts in Japan (in relative isolation from Western European art) would make esthetic judgments showing any degree of similarity to those made by American art experts. Pottery makers seemed a suitable group to study, as the nature of their work calls for some degree of esthetic judgment; yet they are relatively uneducated and isolated and less acquainted with Western traditions of fine art than are many other kinds of artists and collectors in present-day Japan.

B. SUBJECTS

Sixty potters, representing three distinct groups, were interviewed. The largest sample, 36 potters, was from Tanba, a small village where pottery has

* Received in the Editorial Office, Provincetown, Massachusetts, on July 7, 1965, and given special consideration in accordance with our policy for cross-cultural research. Copyright, 1966, by The Journal Press.

¹ The analysis of data and preparation of this report were supported by Cooperative Contract Number 2840 with the U. S. Office of Education.

been made for many centuries. About 35 families are now engaged in pottery making there, each family having its own oven. The second-largest group of 14 potters was from Izushi. In this village there are five ovens, the largest of which has 15 potters at work. The smallest group of subjects, 10 potters, made Kiyomizu ware in the city of Kyoto. Kiyomizu ware is made in large organizations; the one from which we obtained our 10 subjects has over 100 workers.

All the potters work in a definite tradition, a different one for each of the three groups, yet in a manner that permits some degree of individual expression. There is reason to believe, therefore, that experienced and successful potters would have esthetic standards, as well as the purely technical ones required for making replicas of fixed patterns. In all three groups, the potters seem to think of themselves as artists—not as copyists—and are considered as artists by others, but they do not usually sign their work. In general, the potters interviewed have had long experience, averaging 19 years. Only four have had less than five years of experience, while the others have had experience ranging up to 50 years. Most of the persons interviewed began work in their teens. Only four were women. Their educational attainment probably does not exceed the sixth grade, except for the younger Kiyomizu potters who may have been through the ninth. They are not likely to have been exposed to Western art in school nor are there obvious sources of influence from Western art outside school. Japanese newspapers contain few illustrations of art, and the same is true of the magazines likely to be seen by Japanese of the potters' socioeconomic level. Kyoto is a city which many foreigners visit, but even there the potters are not likely to have seen abstract painting deliberately presented as such, even though they may have seen advertising posters influenced by it. It is highly improbable that any of the potters had ever seen any of the works whose reproductions provided the material for this study.

The potters of Izushi and Tanba were interviewed by a university student from that part of Japan who can speak the local dialect. He had previously had training in data collection in social psychology and was prepared for the present work in conferences with Dr. Iwao. The potters in Kyoto were interviewed by Dr. Iwao.

C. MATERIALS JUDGED

Materials judged were of two kinds.

1. The first type consisted of pairs of black-and-white photographs, 4" × 5". Each pair represented two works of art of similar kind, style, and subject matter; and one had been judged better esthetically by the person who had originally picked out the two. Each such pair was first made into slides; and

only those pairs of slides on which at least 12 out of 14 expert judges in New Haven, Connecticut, agreed with the original selector were considered suitable. Prints were then prepared of 51 such pairs considered promising for transcultural study because understanding them seemed to require a minimum of specific cultural knowledge. Since 51 pairs would have taken more time than it was possible for most potters to spare, the set was divided at random into two halves. Twenty-six of the 51 pairs were shown to odd-numbered subjects (numbered, that is, in the order in which they were interviewed); the other 25, to the even-numbered subjects. Meanwhile, the pairs of photographic prints were judged by a new group of New Haven experts; and we required that when shown in this form rather than as projected slides there be no more than three disagreements with the original selector (out of 15 new judges for some of the pairs and 17 for others). Fifteen of the 51 pairs did not meet this requirement of agreement among the New Haven experts; therefore, they were disregarded in the analysis of the Japanese data. It happened that of the remaining pairs whose results were to be analyzed, 21 were among those which had been shown to the odd-numbered subjects and 15 among those shown to the even-numbered subjects.

2. The second type of material consisted of pairs of postcard reproductions (in color) of abstract paintings. Sixteen pairs had been selected on the basis of earlier expert ratings of groups of postcards, and only four of the pairs failed to meet the criterion of no more than three disagreements among 17 new expert judges in New Haven. The 16 pairs of abstract paintings were shown to all 60 Japanese potters, and the judgments on the 12 on which there was excellent agreement in New Haven were analyzed.

D. PROCEDURE

The New Haven judges had been asked to indicate which work of art in each pair was in their opinion better as a work of art—that is, esthetically superior. The Japanese language lends itself to asking substantially the same question. The words used were "*Dochiraga geijutsutekini sugurete iruto omoi masuka?*"

In Izushi and Tanba, there were no barriers to participation in the research. The potters were free to take time off from their work, and a friendly and respectful attitude toward the interviewer as a young man of their locality who had gone to the university in Tokyo favored interest in the project. In Kyoto, on the other hand, with the potters working in something more like a factory system, it was possible to arrange interviews only during lunch hours or other rest periods. This difficulty, however, does not seem likely to bias results; all

the potters approached seemed to be willing to participate, and the limitations came from their working conditions.

Judgments were collected in individual interviews. The pairs were presented one at a time: black-and-white pairs first, then the colored abstracts—and the interviewer recorded which work in each pair was judged to be the better.

E. RESULTS

If the esthetic judgments of the Japanese potters and of the New Haven experts were determined completely independently, one would expect that the odd-numbered potters (who saw 21 black-and-white pairs on which the New Haveners were well agreed) would on the average agree with the New Haven experts on 10.5 pairs. The result was agreement with the New Haveners on 13.1 pairs. The t value for the departure of this mean from the expected value of 10.5 is 5.3. The even-numbered subjects (who saw 15 of the black-and-white pairs on which New Haveners were well agreed) agreed with the New Haveners on an average of 8.8 of the pairs instead of the 7.5 that would be expected on the hypothesis of complete independence. The t for this difference is 3.2. For the colored abstract pairs, all the subjects may be considered together because they saw the same pairs. On the average, they agreed with the New Haven experts on 6.9 of the pairs instead of the six that would be expected on the average by chance. This departure from chance expectation has a t value of 5.8. All these results are significant statistically, leading to a rejection of the null hypothesis at the .01 level or beyond.

Another way of indicating the tendency toward agreement is by explicit reference to individual differences. Of the 30 potters who saw 21 black-and-white pairs on which New Haveners were well agreed, only four failed to agree with the New Haveners on more than half of the pairs. Of the 30 potters who saw 15 black-and-white pairs on which the New Haveners were well agreed, only eight failed to agree with the New Haveners on more than half. The lowest score for agreement was 27 per cent; the highest, 91 per cent. The differences among individual potters in tendency to agree with New Haven experts are not, however, highly consistent. For the group seeing 21 items, the alpha coefficient of internal consistency is only .37; and for the group seeing 15 items, it is .38. For the colored abstract pairs, only five out of the 60 potters agreed with the New Haven experts on fewer than 50 per cent of the pair; another 18 agreed with the New Haveners on exactly half of the pairs; and the remaining 37 agreed with the New Haveners on more than half. The scores range from a low of 33 per cent to a high of 83 per cent. This variation among individuals showed no evidence of internal consistency;

the alpha coefficient was actually negative, a result that could presumably occur only by chance.

The amount of agreement can also be expressed by reference to differences among pairs. Among the 36 black-and-white pairs on which the New Haven judges were well agreed, there were 27 on which a majority of the Japanese potters also agreed; there were eight on which a minority of the potters agreed with the New Haveners; and one on which the potters were split 50-50. The black-and-white pairs varied from one on which only 27 per cent of the potters agreed with the New Haven experts to one pair on which only one out of 30 potters failed to agree with them. An alpha coefficient expressing the average consistency of the ordering of the pairs by a group of 30 potters, with respect to percentage of agreement with New Haven experts, has the value of .69 for the set of 21 black-and-white pairs and .82 for the set of 15, indicating very considerable stability. Among the colored pairs, response varied even more, from 22 per cent to 97 per cent agreement. The alpha coefficient, separately for the two sets of potters shown different sets of black-and-white pairs, is .89 or .86. A majority of the potters agreed with the New Haven experts on only seven out of the 12 pairs. It is because their agreements (on the average) are greater than their disagreements that there is evidence of tendency to agree.

To summarize, both for varied pairs of black-and-white photographs of works of art and for pairs of colored postcards of abstract paintings, there is evidence that Japanese potters tend to agree with American experts in making esthetic judgments, but the amount of agreement is not large. Where complete independence of judgment would be expected to yield 50 per cent agreement, the agreement obtained averages 61 per cent for the black-and-white pairs and 57.5 per cent for the abstract paintings. This deviation from chance agreement is, however, highly significant.

F. DISCUSSION

The most obvious interpretation of our findings, and the one to which the authors are inclined, is to suppose that the data reflect a tendency for people interested in art (regardless of their own particular traditions) to agree on some aspects of esthetic evaluation. On this interpretation, the fact that the agreement between New Haven experts and Japanese potters is not very close may be due to the fact that either or both groups based their evaluations very largely on aspects of the art on which there would be little, if any, tendency toward transcultural agreement.

An alternative to be considered is that the agreement that was found

emerged not from independent currents in Japanese and Western culture, but from the influence of one upon the other. The most conspicuous influence has been that of Western culture upon Japan; thus, the Japanese potters might show some tendency to agree with New Haven experts as a result of some slight or indirect contact with Western art. There is no simple way to deal altogether conclusively with this interpretation. Insofar as it involves a supposition that mere exposure to some Western art would lead anyone to agree with Western experts, we do have some relevant comparisons to make.

Of the 36 black-and-white pairs reported on in this study, 26 were shown (as projected slides) to several hundred suburban-New Haven high-school students, who were asked to express their personal preference between the two works in each pair. Their average agreement with the New Haven experts was 47 per cent, whereas the Japanese potters averaged 63 per cent agreement with the New Haven experts on these particular pairs. A smaller number of these pairs—nine of them—were used in another study in which Connecticut suburban high-school students were asked to make esthetic judgments: i.e., to say which in each pair they considered to be the better work of art. Here the high-school students showed 57 per cent agreement with the New Haven experts instead of the 68 per cent agreement shown by the Japanese potters. In still another instance, 31 of the pairs were shown to Yale College undergraduates, who were asked to make esthetic judgments; they showed 64 per cent agreement with New Haven experts instead of the 61 per cent shown by the Japanese potters. General samples of high-school or college students in or near New Haven, then, agree with New Haven experts either substantially less than the Japanese potters or else to about the same extent, despite their being completely immersed in Western culture and infinitely more experienced with Western art. Clearly, the agreement of the Japanese potters cannot be explained as any simple or automatic consequence of some slight exposure to Western art.

G. SUMMARY

Japanese potters were asked to judge which was the better work of art within each of many pairs of reproductions for which American experts were well agreed about which was better. Thirty potters judged 21 pairs of black-and-white reproductions; another 30 potters, 15 other black-and-white pairs; and all 60 judged 12 pairs of colored abstract paintings. Agreement with the American experts was found in 62 per cent of the judgments on the first batch of black-and-white pairs, 59 per cent of the judgments on the second

batch of black-and-white pairs, and 57.5 per cent of the judgments on the colored pairs. Each of these values differs significantly from the chance value of 50 per cent.

This agreement, because it exceeds that of many nonexpert Americans, must be connected with the artistic interest or activity of the Japanese potters. We propose that its origin lies partly in the independent discovery, by people in differing cultural traditions, of similar facts about the adequacy of particular works for satisfying esthetic interests.

REFERENCES

1. CHILD, I. L., & SIROTO, L. BaKwele and American esthetic evaluations compared. *Ethnology*, 1965, 4, 349-360.
2. FORD, C. S., PROTHRO, E. T., & CHILD, I. L. Some transcultural comparisons of esthetic judgment. *J. Soc. Psychol.*, 1966, 68, 19-26.

Department of Psychology
Yale University
New Haven, Connecticut

Further Evidence of Agreement between
Japanese and American Esthetic Evaluations

Sumiko Iwao

Keio University

Irvin L. Child

Yale University

Miguel García

University of Puerto Rico

In a previous article (Iwao & Child, 1966) we reported that esthetic judgments made by Japanese potters show statistically significant agreement with those made by art students and other experts in the United States. Considered together with other recent evidence (Child & Siroto, 1965; Ford, Prothro, & Child, 1966), this finding suggests that some degree of transcultural agreement in esthetic evaluation may be found in people anywhere who become involved with art, even though preferential reactions by unselected people may not show transcultural agreement. This view conflicts with the prevailing notion that evaluative reactions are purely relative to culture or to personality. Further relevant facts should be useful. We report here on agreement found between American experts and some Japanese involved in art in ways different from the potters.

SUBJECTS

This research was done in Tokyo and its immediate vicinity. We sought to interview people active as teachers or practitioners of at least

Further Evidence of Agreement between
Japanese and American Esthetic Evaluations

Sumiko Iwao

Keio University

Irvin L. Child

Yale University

Miguel García

University of Puerto Rico

In a previous article (Iwao & Child, 1966) we reported that esthetic judgments made by Japanese potters show statistically significant agreement with those made by art students and other experts in the United States. Considered together with other recent evidence (Child & Siroto, 1965; Ford, Prothro, & Child, 1966), this finding suggests that some degree of transcultural agreement in esthetic evaluation may be found in people anywhere who become involved with art, even though preferential reactions by unselected people may not show transcultural agreement. This view conflicts with the prevailing notion that evaluative reactions are purely relative to culture or to personality. Further relevant facts should be useful. We report here on agreement found between American experts and some Japanese involved in art in ways different from the potters.

SUBJECTS

This research was done in Tokyo and its immediate vicinity. We sought to interview people active as teachers or practitioners of at least

one of the traditional Japanese arts. Judgments were obtained from 27 women and 4 men; most were housewives, but a few worked full time at other occupations. All participated, either as an avocation, or a part- or full-time vocation, in one or more of the following activities:

A. Teaching flower arrangement (23 persons). Flower arranging basically follows traditional rules, but the rules alone cannot guarantee a superior outcome. Ability to make sound esthetic judgments seems likely to be essential for either successful practice or successful teaching.

B. Tea ceremony (12 persons). The tea ceremony is probably more bound by convention than is flower arranging; individual style, performance, or interpretation, however, can markedly influence the total effect. The ceremony makes use of objects--tea cup, scroll, and tea room, for example--yet is a temporal art like music or dance and is complex, including movements and utterances as well as objects. Good esthetic judgment seems likely, then, to be valuable for success here too, even if visual judgments in particular play a smaller role than in flower arrangement.

C. Miscellaneous: textile-dyeing (4); doll-making (1); wood-carving (1); Japanese painting (1); calligraphy (1). In addition, several were known to engage in traditional musical or poetic activities.

Our subjects varied in age from 26 to 75 and also varied greatly in amount of artistic experience. Most were friends or teachers of Dr. Iwao's relatives or of her university students.

PROCEDURE

The subjects' cooperation was obtained through personal acquaintance, and they were interviewed individually in either their homes or the homes and workshops of their master teachers. All interviewing was done

in Japanese by native speakers. Dr. Iwao conducted one-third of the interviews, and her undergraduate assistant did the rest.

The subject was asked to indicate which he considered esthetically better in each pair of art works he saw. He was shown a succession of 51 pairs of black-and-white photographs of varied works, and then 24 pairs of colored postcards of modern abstract paintings. Each pair comprised two works similar in kind or subject but differing in esthetic value in the opinion of U. S. experts who had judged them.

Every one of these pairs met the same criteria as did the somewhat smaller number we reported on in the earlier paper about potters. That is, having passed an initial screening which itself was based on considerable agreement among expert judges, the pair had to elicit also a high rate of agreement among a new set of U. S. judges. (The precise criteria are described in the earlier article.)

RESULTS

In response to the 51 varied pairs of black-and-white photographs, the subjects make the same judgment as the majority of U. S. experts 58.5% of the time. This average score for a person's agreement is significantly higher ($t = 4.81$) than the 50% agreement that might on the average be expected in the absence of any bases for consistent agreement or disagreement. Of the 31 individuals, 25 have agreement scores above 50%, reaching as high as 82%; only 6 have agreement scores below 50% (the lowest was 37%). In 36 of the 51 pairs, more than half of our 31 interviewees give the same judgment as the American expert consensus, and in only 15 pairs fewer than half do. On particular pairs the agreement reaches as high as 94% and as low as 26%.

For the colored pairs of abstract paintings, results are very different. These Japanese subjects agree with the American consensus, on the average, only 51.5% of the time. This does not differ significantly from 50%; t is only .80. Individual scores are almost evenly split above and below 50% (14 above, 12 below, 5 exactly at 50%). So are the pairs, with 11 above and 13 below. The agreement with U. S. experts is clearly no more than might arise by chance.

Curiously, even this result for the abstract paintings may indicate some tendency toward the choices made by the American judges. Art pairs selected to present a contrast in esthetic value often pit an esthetic appeal against some other kind of appeal. People lacking an esthetic orientation, then, do not necessarily respond at random; they may systematically choose on the basis of these non-esthetic appeals. The tendency is very clearly seen in the pairs put together by Margaret Bulley (1951) that suggested the technique we are using. We deliberately attempted to make non-esthetic appeal more equal for the two works in a pair, but were not uniformly successful. Some evidence, however, described below, enables us to take account of the non-esthetic appeal in arriving at an interpretation of our results.

What we need is information about responses to our pairs by a group in which esthetic orientation is likely to be rare. Will the choices made in such a group agree with expert evaluation about 50% of the time? Or will the choices systematically disagree with expert evaluation? Most of our black-and-white pairs have been shown as projected slides to groups of school children, including some groups in which an esthetic orientation is rare. But our colored pairs have not been, and in any event responses made by adults, and to stimuli in the same form, would be desirable.

Fortunately, both kinds of our pairs, in exactly the same form we used in Tokyo, have been shown to one adult group in which esthetic orientation seems likely to be rare. This is a sample of 40 working-class residents of Puerto Rico, whose preferences were obtained by Miguel García. These people, like our Japanese subjects, were interviewed in their own language by a native speaker; they were, however, asked about their personal preference, not about esthetic merit. On the 51 black-and-white pairs, the preferences they expressed agreed with the U. S. judgments only 43% of the time--a highly significant tendency toward disagreement. On the 24 colored abstract pairs, their average agreement was 45.5%, again significantly below 50%. The difference between this and the 51.5% agreement of our Japanese subjects is highly significant ($t = 2.92$).

Thus, in response to the abstract paintings our Japanese subjects, while not showing an absolute tendency to agree with U. S. experts, may well be exhibiting a relative tendency. They certainly are not responding at random; the very great difference between their response to one pair and to another enables us to confidently reject that possibility. The most reasonable inference is that they share some tendency with the U. S. experts to a sufficient degree to keep them from exhibiting a contrary preference manifested by at least some groups lacking an esthetic orientation.

DISCUSSION

These findings add to the evidence that esthetic evaluations developed in two distinct cultural traditions have some tendency to agree. The American judges here are the same who entered our previous comparison with Japanese potters. The Japanese judges in the present study, living in Tokyo, are not so isolated from Western influence as were the majority of

the potters. Yet their artistic involvement is decidedly in Japanese tradition, and it seems likely that their esthetic evaluations originate primarily in it (insofar as they are to be considered cultural rather than individual). Quite directly for most of our stimuli, and by an indirect comparison in the case of the abstract paintings, we find evidence that their esthetic evaluations, like those of the potters, tend toward agreement with those by Americans involved with art.

Our findings also argue rather persuasively against the explanation of transcultural agreement as entirely a product of diffusion. The Japanese subjects in the present study have certainly been exposed to possible influences from Western art. Some of them, for example, said they had seen exhibits of Western art--an experience that, almost certainly, none of the potters had ever had. Yet instead of agreeing with American judges more than did the potters, they agreed somewhat less.

SUMMARY

Thirty-one residents of Tokyo, all practitioners or teachers of flower arranging, tea ceremony, or other traditional arts, judged which was the esthetically better work in each of 51 pairs of black-and-white photographic prints and 24 pairs of colored abstract-painting postcards. U. S. experts had shown good agreement among themselves about which work in each pair was better. The Japanese subjects showed significant tendency to agree with the U. S. experts on the prints. They did not on the abstract paintings; yet comparison with a group of adults lacking esthetic orientation showed these Japanese to differ significantly from the decided tendency of the other group to prefer the works spurned by U. S. experts. These findings add to the growing evidence that esthetic evaluations developed in different cultural settings may tend toward agreement.

REFERENCES

- Bulley, Margaret H. Art and everyman. Vol. 1. London: Batsford, 1951.
- Child, I. L., & Siroto, L. BaKwele and American esthetic evaluations compared. Ethnology, 1965, 4, No. 4, 349-360.
- Ford, C. S., Prothro, E. T., & Child, I. L. Some transcultural comparisons of esthetic judgment. Journal of Social Psychology, 1966, 68, 19-26.
- Iwao, Sumiko, & Child, I. L. Comparison of esthetic judgments by American experts and by Japanese potters. Journal of Social Psychology, 1966, 68, 27-33.

Personality and Esthetic Sensitivity: Extension of Findings
to Younger Age and to Different Culture

Irvin L. Child
Yale University
and Sumiko Iwao
Keio University

Esthetic sensitivity, measured by the extent to which a person's judgments or preferences about visual art agree with those of experts, has recently been found related to personality variables indicating cognitive independence and openness (Child, 1962; 1965). These findings challenge the common assumption that esthetic evaluation is based principally on passive acceptance of convention, and their validity is therefore crucial to psychological interpretation of reactions to art.

The notion that esthetic sensitivity is related to cognitive independence and openness has thus far, however, a very narrow evidential base. While evidence comes from several separate studies and is highly significant statistically, it pertains entirely to undergraduate men at two American universities of similar type. This population is selected in a variety of ways which could conceivably create correlations not to be found elsewhere. The correlations might then be of very slight interest, applying only to a limited population; or perhaps of no interest, being an artifact of selection.

We report here two tests of the generality of this finding. First, it is tested for a younger group, more diversified in academic ability, and evenly divided between the sexes, but in the same country. Second, it is tested for a

group like the original one in sex and social status--college undergraduates--but in a country of greatly different culture, Japan.

Study of Secondary-School Students

In a survey of school children's art preferences, a measure of esthetic sensitivity was obtained on nearly every pupil in the secondary schools (grades 7 through 12) of a suburban New England town with a diversified population. The materials consisted, as in the college study, of pairs of slides showing works of art; the two pictures in a pair were similar in kind or subject matter and usually in style, but one was superior esthetically in the opinion of expert judges. Secondary-school students were shown these pairs without being told their origin, and were asked to indicate personal preference within each pair; the research was described to them as a survey of art preferences. This was indeed its main purpose. Different groups of students saw different sets of pairs; in all, we used 7 different sets of 130 pairs each. Each student could be scored for esthetic sensitivity by counting the number of his preferences, out of 130, which agreed with the esthetic judgments of experts.

While the survey of preferences--having a direct educational justification--utilized the entire school population, the exploration of personality variables had to be done with selected samples only and with an instrument requiring little time. For each grade from 7 through 12, we picked from the data on esthetic sensitivity the three highest-scoring boys and the three highest-scoring girls. Their names were submitted to the school guidance staff along with the names of 10 or 12 of the lowest scorers of each sex and grade. A member of the guidance staff then selected from the latter pool the individual closest to each of the high scorers in general academic potential. This matching was done to control the difference in general academic potential which

might well accompany the difference between high and low scorers, so that it could neither drown out nor produce personality differences. Those doing the matching were not familiar with the hypotheses guiding the research.

From the parents of children originally selected in this way, permission was requested by letter (with telephone follow-up if needed) to interview their child if the child was also willing. The basis for selection of children was not revealed. Permission was refused for one of the high scorers and five of the low scorers--in at least two instances at the initiative of the child. Four others could not be interviewed for other reasons (difficulty in scheduling, or transfer to another school). Substitutes were chosen for these ten so as to preserve the pattern of matching.

The interview consisted partly of an oral inquiry into reasons for preferences; this was part of another research program which will be reported on later. The rest was a personality questionnaire which the student filled out, made up of items drawn from the questionnaires previously used with college students but in many instances re-worded to make them more understandable and more relevant to children as young as 7th-graders. (This questionnaire is reproduced in Appendix B of Child, 1964.)

The questionnaire measured three variables which had most consistently shown a relation to esthetic sensitivity in the prior research on college students: (1) Independence of Judgment, a scale derived from Barron's measure by this name (Barron, 1953) but here represented by 19 items of which 12 had been altered for use with younger subjects. (2) Tolerance of Complexity, a scale described in previous research by Child (1965) and based partly on items from the F-Test of Adorno et al. (1950); here represented by 13 items of which 11 had been altered. (3) Regression in the Service of the Ego, a questionnaire developed by David Singer and represented here by 16 items, all of which had

been altered. Some changes were slight; a few involved complete re-wording. These scales would not be very satisfactory for drawing any conclusion about an individual, for the internal consistency is not high. The alpha coefficient for the 36 high scorers and for the 36 low scorers in turn is, for Independence of Judgment .45 and .39; for Tolerance of Complexity .52 and .28; and for Regression in the Service of the Ego .53 and .42. But these values for internal consistency are high enough to give some hope that the scales could be useful in discriminating among groups.

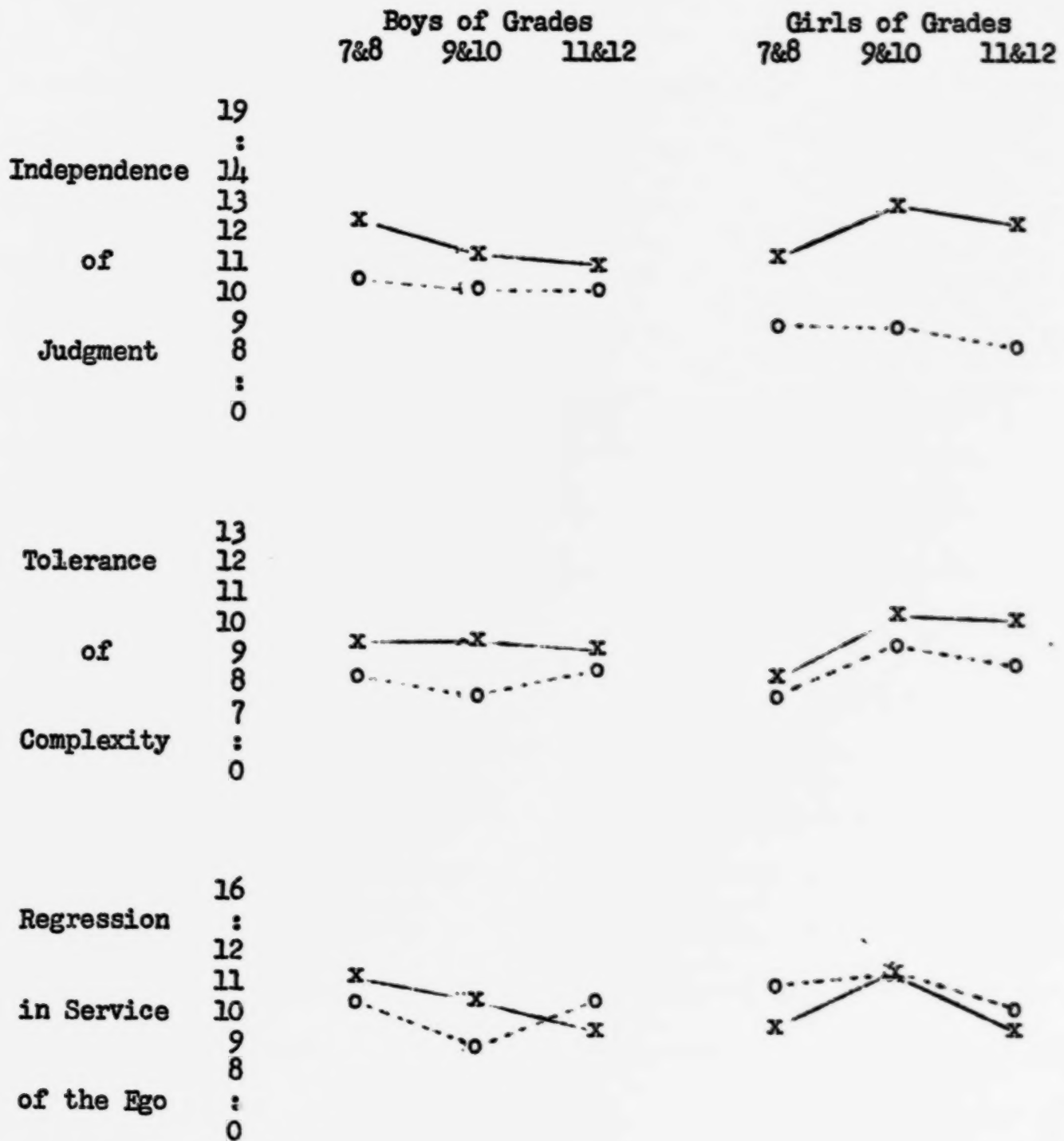
The results are given in Figure 1. In preparing the graphs we have pooled adjacent grades so that each point could be based on six subjects rather than three; in doing analyses of variance to determine the significance of findings, we have kept each grade separate.

For two of the personality variables, Independence of Judgment and Tolerance of Complexity, there are consistent differences between high and low scorers in the same direction found in the college studies. For Independence of Judgment the over-all difference is significant at the .001 level (F of 17.3 with 1 and 46 degrees of freedom); for Tolerance of Complexity it is significant at the .01 level (F of 6.2 with 1 and 46 degrees of freedom, one-tailed test). Only one interaction of these measures with sex or grade even approaches being significant: it is the tendency for Independence of Judgment to distinguish between high and low scorers more among girls than among boys, which reaches the 10% level of significance but not the 5% (F of 3.2 with 1 and 46 degrees of freedom, two-tailed test). For these two variables, then, the college results are confirmed for a more heterogeneous population at a lower average intellectual level, and for both sexes.

For the third variable, Regression in the Service of the Ego, there are no consistent findings, and the college results are not confirmed. The college

Figure 1
Personality Questionnaire Scores: Comparison of Secondary-School
Students High and Low in Esthetic Preference

(Each entry is the mean score of 6 students.)



Key:
 x = Students high in esthetic preference
 o = " low " " "

results were so stable that they are not in doubt for the particular population studied. Thus we may have discovered here a real difference in what personality variables are important for esthetic sensitivity at different socio-cultural levels or at different ages. Or the findings either at college or at secondary level may turn out to be unrepresentative for a larger population, to have resulted somehow from peculiarities of the specific groups studied.

Since the internal consistency of the personality variables was not very high, it may be of interest to ask about the number of separate items on which high and low scorers are differentiated in the expected direction. Of the Independence of Judgment items, 14 show a difference in the expected direction and only 3 show an opposite difference. Of the Tolerance of Complexity items, 10 give the expected difference and 3 the opposite. The Regression in the Service of the Ego items, however, are evenly divided; 8 show the expected difference and 8 the opposite. Thus the item results lead to the same conclusion as does the scoring of the variables: most significant confirmation of prediction for Independence of Judgment, next for Tolerance of Complexity, and none for Regression in the Service of the Ego.

Study of College Men in Japan

The second study was of 131 male undergraduates at Keio University, located in Tokyo and drawing its students from all parts of Japan and from a variety of social backgrounds. The subjects were obtained through classes and extra-curricular activities--journalism club, hiking club, etc.; sometimes all present at a meeting took part right then, while in other instances volunteers were solicited for a later appointment. Some of the groups included a few women, and they took part also; but since women totaled only 10 in our two conditions combined, we were not able to use the data from them.

The stimulus materials were 80 pairs of slides drawn from those used in the U. S. studies. They were the same 80 used in the partial replication of the U. S. college study described by Child (1965, pp. 502-506), and were presented in the same order. They had been selected for trans-cultural use as requiring little knowledge of any specific tradition for understanding the work or subject matter. As in the U. S. replication, instructions given to about half of the subjects called for an esthetic judgment about each pair, and the others were asked to express personal preferences.

The questionnaire consisted of Japanese versions of items which in the U. S. replication had yielded most consistent relations with esthetic sensitivity, and which were intended to measure either (1)Independence of Judgment, (2)Regression in the Service of the Ego, or (3)Tolerance of Complexity, Ambivalence, Ambiguity, and Unrealistic Experience. The first two of these are among the variables that had yielded most consistent results and had therefore been included in the secondary-school study described above. The third includes the third variable of the secondary-school study but embraces also items from related variables for which as a whole earlier results had not been so significant. Specifically, the Japanese version included all items for these variables which, in the U. S. college replication, had shown separate item correlations of at least $+.10$ both with the judgment measure and with the preference measure of esthetic sensitivity, except that one item meeting this criterion was omitted because of Japanese unfamiliarity with the subject matter (sky diving).

The internal consistency of the three variables is distressingly low, with alpha coefficients--shown in Table 1--varying from $.20$ to $.36$. For two of the scales, however, these coefficients are not lower than the U. S. results by more than one might expect from the fact that fewer items are being used. Only for the Tolerance variable is the coefficient much lower, and this is the

variable which was more broadly defined for the Japanese questionnaire. There seems no reason to suppose that the personality items we are using have really any lower consistency in their Japanese than in their American use; the variables are less reliably measured because of the smaller number of items and the greater diversity of one scale.

Relation of these three variables to esthetic sensitivity is given in Table 1, separately for the 66 judgment subjects and the 65 preference subjects. Since the results look very similar for the two groups, they are then presented pooled (the pooling was done by averaging the two correlations). Every correlation is in the predicted direction. The pooled results are significant for each variable. The outcome is similar to that for the American secondary-school students in being least strong for Regression in the Service of the Ego. But it differs in being confirmatory even for this variable, and in providing on the Tolerance variable evidence at least as strong as that on Independence of Judgment.

Again the results may be considered item-by-item. For each item, two correlations with esthetic sensitivity are available, one from the judgment subjects and one from the preference subjects. For Independence of Judgment, both correlations are positive for 6 items, both are negative for only one item; for each of the other two items, the correlations are split, one positive and one negative. For Tolerance, the correlations are both positive for 4 items and are split for 7 items; there are no items with both negative. For Regression in the Service of the Ego, both are positive for 2 items and both are negative for one item; the other 3 items have split results. Again, the confirmation is least for Regression in the Service of the Ego.

With these Japanese data, since they are not based on extreme groups alone, we may apply the technique of multiple correlation to measure to what

TABLE 1

Personality Variables in Japanese College Students:
Internal Consistency and Relation to Esthetic Sensitivity

(In parentheses are given the most nearly comparable correlations reported by Child, 1965 for American college students.)

Variable	No. of items	Alpha coefficient of internal consistency	Correlation with esthetic judgment	Correlation with esthetic preference	Pooled correlations
		N = 131	N = 66	N = 65	N = 131
Independence of Judgment	9	.34	+.14 (.26)	+.26* (.32)	+.20*
Tolerance of Complexity, Unrealistic Experience, Ambiguity, and Ambivalence	11	.20	+.20* (.24)	+.26* (.21)	+.23**
Regression in the Service of the Ego	6	.36	+.19 (.31)	+.12 (.16)	+.16*

* = Significant at the 5% level (one-tailed test)

** = Significant at the 1% level (one-tailed test)

extent esthetic sensitivity is predictable from the three personality variables together. The outcome is a multiple correlation of $+.28$ for esthetic judgment and $+.35$ for esthetic preference. The first of these is decidedly lower than the corresponding value of $+.45$ for the most nearly comparable U. S. data which have been analyzed, those of 84 subjects in the replication reported in Child (1965, pp. 502-506). The second multiple correlation for the Japanese data, however, is of about the same size as the comparable value for 88 U. S. subjects, $+.34$.

University students in Japan are of course subject to a high degree of Westernization, and it seemed possible that degree of familiarity with Western art might be an important influence on extent of agreement with Western judges on esthetic evaluation. In turn, the personality traits being assessed might conceivably be characteristically Western, and a correlation between them and esthetic sensitivity might thus be a product simply of general degree of Westernization of the individual. We therefore had each subject rate the extent of his familiarity with Western art. This rating did not have with esthetic sensitivity the large positive correlation we feared it might. The tendency was in this direction, but it was not statistically significant: a correlation of $.07$ with esthetic judgment and $.16$ with esthetic preference. The personality variables also showed uniformly positive correlations with self-rating of familiarity with Western art, but with one exception these too were small and not statistically significant; the averaged correlation for judgment and preference subjects was $.13$ for Independence of Judgment, $.18$ for Tolerance, and $.34$ for Regression in the Service of the Ego. These correlations are not of a magnitude to argue for familiarity with Western art as the underlying variable influencing all the others. On the other hand, our hypothesis that esthetic sensitivity develops from cognitive independence and freedom would lead to an expectation

that interest in Western (as well as Oriental) art might well accompany it.

Summary and Discussion

We have reported here two confirmations of a positive relationship, previously found in U. S. college men, of esthetic sensitivity to cognitive independence and openness. A study of U. S. secondary-school students confirms the general relationship for both sexes, a younger age, and a more diversified population, while not confirming a specific relationship with a measure of Regression in the Service of the Ego. A study of college men in Japan confirms the relationship for all measures used.

What should we conclude from the trans-cultural confirmation? Subsidiary evidence seems to justify our rejecting the hypothesis that it is an artifact, a common effect of degree of Westernization on response to art and to questionnaire items. We would tentatively conclude that perhaps this relation between personality and esthetic sensitivity may be found in any society where esthetic values are stressed in some generally available part of the cultural tradition--as in Japan and in Western European tradition--so that the individual with cognitive independence and openness has esthetic values available to him as one possible medium for expression and gratification of these cognitive tendencies.

References

- Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. The authoritarian personality. New York: Harper, 1950.
- Barron, F. Some personality correlates of independence of judgment. Journal of Personality, 1953, 21, 287-297.
- Child, I. L. Personal preferences as an expression of aesthetic sensitivity. Journal of Personality, 1962, 30, 496-512.

Child, I. L. Development of sensitivity to esthetic values. Cooperative Research Project No. 1748, Cooperative Research Program of U. S. Office of Education, 1964 (mimeographed).

Child, I. L. Personality correlates of esthetic judgment in college students. Journal of Personality, 1965, 33, 476-511.